

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: August 23, 2004, 14:31:21 ; Search time 12 Seconds

(without alignments)
1067.438 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357

Sequence: 1 MADNFSHDLALSGSNPNQ.....ISKLGISGIDIDITSASYTMI 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1310	96.5	249	1	LEGG3_HUMAN
2	1091	80.4	295	1	P34486 canis famill
3	1077	79.4	263	1	LEGG3_MOUSE
4	1074	79.1	261	1	P08699 rattus norv
5	1044.5	77.0	244	1	LEGG3_CRIO
6	1019	75.1	241	1	LEGG3_RABIT
7	299	22.0	353	1	LEGG9_MOUSE
8	289	21.3	355	1	LEGG9_HUMAN
9	276.5	20.4	354	1	LEGG9_RAT
10	254	18.7	323	1	LEGG4_PIG
11	250.5	18.5	344	1	LEGG5_RAT
12	245	18.1	301	1	LEGG6_MOUSE
13	236.5	17.4	324	1	LEGG4_RAT
14	228.5	16.8	316	1	LEGG8_HUMAN
15	218.5	16.1	316	1	LEGG8_RAT
16	215.5	15.9	323	1	LEGG4_HUMAN
17	211	15.3	135	1	LEGG7_HUMAN
18	207.5	15.3	316	1	LEGG8_MOUSE
19	198	14.6	245	1	CM32_HUMAN
20	197.5	14.6	242	1	CM32_MOUSE
21	196.5	14.5	466	1	ANX7_HUMAN
22	193	14.2	1049	1	ANX7_HUMAN
23	192.5	14.2	463	1	ANX7_MOUSE
24	188.5	13.9	279	1	LEGG3_MOUSE
25	188	13.9	467	1	CBPA_DICDI
26	187.5	13.8	283	1	LEGG1_HABCO
27	187	13.8	684	1	CA39_HUMAN
28	185	13.6	744	1	CA18_HUMAN
29	184	13.6	744	1	LEGG7_RAT
30	183.5	13.5	743	1	LEGG7_MOUSE
31	181	13.3	336	1	LEGGC_HUMAN
32	180	13.3	744	1	CA18_RABIT
33	179	13.2	316	1	CC12_MOUSE

34	179	13.2	316	1	CC13_MOUSE	P20631 caenorhabdi
35	174.5	12.9	296	1	CC01_MOUSE	P08124 caenorhabdi
36	174.5	12.9	299	1	CC34_MOUSE	P34687 caenorhabdi
37	173.5	12.8	266	1	YXWK_MOUSE	O21184 caenorhabdi
38	173.5	12.8	512	1	ANX7_XENLA	O92125 xenopus lae
39	173.5	12.8	754	1	CA54_MOUSE	O28247 canis fam
40	173	12.7	675	1	LEGG7_MOUSE	O54974 mus musculu
41	173	12.7	675	1	CA39_CHICK	P35017 gallus gall
42	172.5	12.7	1464	1	CA13_MOUSE	P08121 mus musculu
43	172	12.7	301	1	CC02_MOUSE	P17656 caenorhabdi
44	171.5	12.6	464	1	S3A2_MOUSE	O15428 homo sapien
45	171.5	12.6	1712	1	CA24_HUMAN	P08572 homo sapien
46	170	12.5	1466	1	CA13_HUMAN	P02461 homo sapien
47	169.5	12.5	302	1	CCDC_MOUSE	P17657 caenorhabdi
48	169.5	12.5	377	1	PEB1_MOUSE	O94615 drosophila
49	169.5	12.5	505	1	ANXB_MOUSE	P50995 homo sapien
50	169	12.5	1669	1	CA14_HUMAN	P02462 homo sapien
51	168.5	12.4	1758	1	CA24_MOUSE	P17140 caenorhabdi
52	168	12.4	635	1	CA28_HUMAN	P25067 homo sapien
53	166.5	12.3	306	1	CC40_MOUSE	P34804 caenorhabdi
54	166.5	12.3	309	1	CC40_MOUSE	P17277 gallus gall
55	166.5	12.3	503	1	ANXB_MOUSE	P34377 oryctolagus
56	166.5	12.3	1262	1	CA13_MOUSE	P12105 gallus gall
57	165	12.2	210	1	CA2C_MOUSE	P16252 haemochus
58	165	12.2	297	1	LEGG3_MOUSE	O09581 caenorhabdi
59	165	12.2	3124	1	CA1C_MOUSE	P13944 gallus gall
60	164.5	12.1	1685	1	CA54_HUMAN	P23400 homo sapien
61	163.5	12.0	263	1	Y737_MOUSE	P83475 drosophila
62	163	12.0	864	1	EUS_MOUSE	O99372 rattus norv
63	163	12.0	1453	1	CA11_MOUSE	P02452 homo sapien
64	163	12.0	1464	1	CA11_MOUSE	O62203 mus musculu
65	162.5	12.0	475	1	S3A2_MOUSE	

ALIGNMENTS

RESULT 1	ID	LEGG3_HUMAN	STANDARD	PRT:	249 AA.
AC	P17931	Q16005; 196647;			
DT	01-NOV-1990	(Rel. 16, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	Galactin-3 (galactose-specific lectin 3) (MAC-2 antigen) (IGF-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (GBP 35)				
DE	(Lectin-binding protein) (Lectin L-29) (L-31) (Galactoside-binding protein) (GALBP)				
GN	LGALS3 OR MAC2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91084480; PubMed=2261464;				
RT	Robertson M.W., Albrandt K., Keller D., Liu F.-T.;				
RT	"Human IGF-binding protein: a soluble lectin exhibiting a highly				
RT	conserved interspecies sequence and differential recognition of IGF				
RT	glycoforms."				
RL	Biochemistry 29:8093-8100(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Cartiloma;				
RX	MEDLINE=90364999; PubMed=2402511.				
RA	Cherayil B., Chaitovitz S., Wong C., Pillai S.;				
RT	"Molecular cloning of a human macrophage lectin specific for				
RT	galactose."				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7324-7326(1990).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91216471; PubMed=2022338;				
RA	Oda Y., Leftler H., Sakakura Y., Kasai K.I., Baronides S.H.;				

RT "Human breast carcinoma cDNA encoding a galactoside-binding lectin
 RT homologous to mouse Mac-2 antigen.";
 RL Gene 99:279-283 (1991).
 CC [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=91183475; PubMed=2009535;
 RA Raz A., Cami P., Raz T., Hogan V., Mohamed A., Wolman S.R.;
 RT "Molecular cloning and chromosomal mapping of a human galactoside-
 RT binding protein.";
 RL Cancer Res. 51:2173-2178 (1991).
 CC [5]
 RN SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
 RX MEDLINE=93234518; PubMed=7682704;
 RA Lotz M.M., Andrews C.W. Jr., Korzeilus C.A., Lee E.C.,
 RA Steele G.D. Jr., Clarke A., Mercutio A.M.;
 RT "Decreased expression of Mac-2 (carbohydrate binding protein 35) and
 RT loss of its nuclear localization are associated with the neoplastic
 RT progression of colon carcinoma.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3466-3470 (1993).
 CC [6]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Gastric adenocarcinoma;
 RA Kato S.;
 RT "Human galectin-3 full-length cDNA.";
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC [7]
 RN SEQUENCE FROM N.A.
 RA Kadofsky M.W., Wang J.L.;
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 CC [8]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Altschul S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Manusia K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.V., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bask S.A., McKwan P.J., McKernan K.J., Malek J.A., Guntaratne P.H.,
 RA Richards S., Morley K.C., Hale S.J., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Murthy K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Tuchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC [9]
 RN PHOSPHORYLATION.
 RX MEDLINE=94075369; PubMed=8253806;
 RA Huflejt M.E., Turck C.W., Lindstedt R., Barondes S.H., Lefler H.;
 RT "L-29, a soluble lactose-binding lectin, is phosphorylated on serine
 RT 6 and serine 12 in vivo and by casein kinase I.";
 RL J. Biol. Chem. 268:26712-26718 (1993).
 CC [10]
 RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 113-249.
 RX MEDLINE=98250755; PubMed=9582341;
 RA Seetharaman J., Karisberg A., Slaaby R., Lefler H., Barondes S.H.,
 RA Rind J.M.;
 RT "X-ray crystal structure of the human galectin-3 carbohydrate
 RT recognition domain at 2.1-A resolution.";
 RL J. Biol. Chem. 273:13047-13052 (1998).
 CC -I- FUNCTION: Galactose-specific lectin which binds IgE.
 CC -I- SUBUNIT: Probably forms homo- or heterodimers.
 CC -I- SUBCELLULAR LOCATION: Nuclear. Cytoplasmic in adenomas and
 CC carcinomas.
 CC -I- TISSUE SPECIFICITY: A major expression is found in the colonic

CC epithelium. It is also abundant in the activated macrophages.
 CC -I- SIMILARITY: In the C-terminal section; belongs to the galectin
 CC (galactin/S-lectin) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M57710; AAA5607.1; -;
 CC EMBL: M35368; AAA8086.1; -;
 CC EMBL: M35682; AAA36163.1; -;
 CC EMBL: M64303; -; NOT_ANNOTATED_CDS.
 CC EMBL: S59012; AA52629.1; -;
 CC EMBL: AB006780; BAA22164.1; -;
 CC EMBL: AF031425; AAB86584.1; -;
 CC EMBL: AF031422; AAB86584.1; JOINED.
 CC EMBL: AF031423; AAB86584.1; JOINED.
 CC EMBL: AF031424; AAB86584.1; JOINED.
 CC EMBL: BC001120; AAH01120.1; -;
 CC EMBL: BC053667; AAH53667.1; -;
 CC PIR: A35820; A35820.
 CC PDB: 1A3K; 15-JUL-98.
 CC Genew: HGNC:6563; LGALS3.
 CC MIM: 153619; -;
 CC GO: GO:0005866; C:plasma membrane, TAS.
 CC GO: GO:0005530; F:lectin, TAS.
 CC InterPro: IPR008985; Cona like _lec_gl.
 CC InterPro: IPR001079; Galectin.
 CC Pfam: PF00337; Gal-bind_lectin; 1.
 CC SMART: SM00276; GLECT; 1.
 CC PROSITE: PS00309; GALAPTIN; 1.
 CC Galectin; Lectin; IgE-binding protein; Repeat; Phosphorylation;
 CC Acetylation; Nuclear protein; Polymorphism; 3D-structure.
 CC KW ACETYLATION; NUCLEAR PROTEIN; POLYMORPHISM; 3D-STRUCTURE;
 CC REPEAT; LECTIN; IGE-BINDING PROTEIN; PHOSPHORYLATION;
 CC ACETYLATION; POLYMORPHISM; 3D-STRUCTURE;
 CC BY SIMILARITY.
 CC ACETYLATION (BY SIMILARITY).
 CC PHOSPHORYLATION (PARTIAL).
 CC PHOSPHORYLATION (REPEAT).
 CC 8 X 9 AA TANDEN REPEATS OF Y-P-G-X(3)-P-
 CC G-A.
 CC 1. REPEAT 35 43
 CC 2. REPEAT 44 52
 CC 3. REPEAT 53 61
 CC 4 (APPROXIMATE). 68
 CC 5 (APPROXIMATE). 77
 CC 6 (APPROXIMATE). 87
 CC 7 (APPROXIMATE). 99
 CC 8 (APPROXIMATE). 108
 CC 9 (APPROXIMATE). 117
 CC 10 (APPROXIMATE). 124
 CC 11 (APPROXIMATE). 132
 CC 12 (APPROXIMATE). 140
 CC 13 (APPROXIMATE). 148
 CC 14 (APPROXIMATE). 156
 CC 15 (APPROXIMATE). 164
 CC 16 (APPROXIMATE). 172
 CC 17 (APPROXIMATE). 180
 CC 18 (APPROXIMATE). 186
 CC 19 (APPROXIMATE). 194
 CC 20 (APPROXIMATE). 202
 CC 21 (APPROXIMATE). 210
 CC 22 (APPROXIMATE). 218
 CC 23 (APPROXIMATE). 226
 CC 24 (APPROXIMATE). 234
 CC 25 (APPROXIMATE). 242
 CC 26 (APPROXIMATE). 250
 CC 27 (APPROXIMATE). 258
 CC 28 (APPROXIMATE). 266
 CC 29 (APPROXIMATE). 274
 CC 30 (APPROXIMATE). 282
 CC 31 (APPROXIMATE). 290
 CC 32 (APPROXIMATE). 298
 CC 33 (APPROXIMATE). 306
 CC 34 (APPROXIMATE). 314
 CC 35 (APPROXIMATE). 322
 CC 36 (APPROXIMATE). 330
 CC 37 (APPROXIMATE). 338
 CC 38 (APPROXIMATE). 346
 CC 39 (APPROXIMATE). 354
 CC 40 (APPROXIMATE). 362
 CC 41 (APPROXIMATE). 370
 CC 42 (APPROXIMATE). 378
 CC 43 (APPROXIMATE). 386
 CC 44 (APPROXIMATE). 394
 CC 45 (APPROXIMATE). 402
 CC 46 (APPROXIMATE). 410
 CC 47 (APPROXIMATE). 418
 CC 48 (APPROXIMATE). 426
 CC 49 (APPROXIMATE). 434
 CC 50 (APPROXIMATE). 442
 CC 51 (APPROXIMATE). 450
 CC 52 (APPROXIMATE). 458
 CC 53 (APPROXIMATE). 466
 CC 54 (APPROXIMATE). 474
 CC 55 (APPROXIMATE). 482
 CC 56 (APPROXIMATE). 490
 CC 57 (APPROXIMATE). 498
 CC 58 (APPROXIMATE). 506
 CC 59 (APPROXIMATE). 514
 CC 60 (APPROXIMATE). 522
 CC 61 (APPROXIMATE). 530
 CC 62 (APPROXIMATE). 538
 CC 63 (APPROXIMATE). 546
 CC 64 (APPROXIMATE). 554
 CC 65 (APPROXIMATE). 562
 CC 66 (APPROXIMATE). 570
 CC 67 (APPROXIMATE). 578
 CC 68 (APPROXIMATE). 586
 CC 69 (APPROXIMATE). 594
 CC 70 (APPROXIMATE). 602
 CC 71 (APPROXIMATE). 610
 CC 72 (APPROXIMATE). 618
 CC 73 (APPROXIMATE). 626
 CC 74 (APPROXIMATE). 634
 CC 75 (APPROXIMATE). 642
 CC 76 (APPROXIMATE). 650
 CC 77 (APPROXIMATE). 658
 CC 78 (APPROXIMATE). 666
 CC 79 (APPROXIMATE). 674
 CC 80 (APPROXIMATE). 682
 CC 81 (APPROXIMATE). 690
 CC 82 (APPROXIMATE). 698
 CC 83 (APPROXIMATE). 706
 CC 84 (APPROXIMATE). 714
 CC 85 (APPROXIMATE). 722
 CC 86 (APPROXIMATE). 730
 CC 87 (APPROXIMATE). 738
 CC 88 (APPROXIMATE). 746
 CC 89 (APPROXIMATE). 754
 CC 90 (APPROXIMATE). 762
 CC 91 (APPROXIMATE). 770
 CC 92 (APPROXIMATE). 778
 CC 93 (APPROXIMATE). 786
 CC 94 (APPROXIMATE). 794
 CC 95 (APPROXIMATE). 802
 CC 96 (APPROXIMATE). 810
 CC 97 (APPROXIMATE). 818
 CC 98 (APPROXIMATE). 826
 CC 99 (APPROXIMATE). 834
 CC 100 (APPROXIMATE). 842
 CC 101 (APPROXIMATE). 850
 CC 102 (APPROXIMATE). 858
 CC 103 (APPROXIMATE). 866
 CC 104 (APPROXIMATE). 874
 CC 105 (APPROXIMATE). 882
 CC 106 (APPROXIMATE). 890
 CC 107 (APPROXIMATE). 898
 CC 108 (APPROXIMATE). 906
 CC 109 (APPROXIMATE). 914
 CC 110 (APPROXIMATE). 922
 CC 111 (APPROXIMATE). 930
 CC 112 (APPROXIMATE). 938
 CC 113 (APPROXIMATE). 946
 CC 114 (APPROXIMATE). 954
 CC 115 (APPROXIMATE). 962
 CC 116 (APPROXIMATE). 970
 CC 117 (APPROXIMATE). 978
 CC 118 (APPROXIMATE). 986
 CC 119 (APPROXIMATE). 994
 CC 120 (APPROXIMATE). 1002
 CC 121 (APPROXIMATE). 1010
 CC 122 (APPROXIMATE). 1018
 CC 123 (APPROXIMATE). 1026
 CC 124 (APPROXIMATE). 1034
 CC 125 (APPROXIMATE). 1042
 CC 126 (APPROXIMATE). 1050
 CC 127 (APPROXIMATE). 1058
 CC 128 (APPROXIMATE). 1066
 CC 129 (APPROXIMATE). 1074
 CC 130 (APPROXIMATE). 1082
 CC 131 (APPROXIMATE). 1090
 CC 132 (APPROXIMATE). 1098
 CC 133 (APPROXIMATE). 1106
 CC 134 (APPROXIMATE). 1114
 CC 135 (APPROXIMATE). 1122
 CC 136 (APPROXIMATE). 1130
 CC 137 (APPROXIMATE). 1138
 CC 138 (APPROXIMATE). 1146
 CC 139 (APPROXIMATE). 1154
 CC 140 (APPROXIMATE). 1162
 CC 141 (APPROXIMATE). 1170
 CC 142 (APPROXIMATE). 1178
 CC 143 (APPROXIMATE). 1186
 CC 144 (APPROXIMATE). 1194
 CC 145 (APPROXIMATE). 1202
 CC 146 (APPROXIMATE). 1210
 CC 147 (APPROXIMATE). 1218
 CC 148 (APPROXIMATE). 1226
 CC 149 (APPROXIMATE). 1234
 CC 150 (APPROXIMATE). 1242
 CC 151 (APPROXIMATE). 1250
 CC 152 (APPROXIMATE). 1258
 CC 153 (APPROXIMATE). 1266
 CC 154 (APPROXIMATE). 1274
 CC 155 (APPROXIMATE). 1282
 CC 156 (APPROXIMATE). 1290
 CC 157 (APPROXIMATE). 1298
 CC 158 (APPROXIMATE). 1306
 CC 159 (APPROXIMATE). 1314
 CC 160 (APPROXIMATE). 1322
 CC 161 (APPROXIMATE). 1330
 CC 162 (APPROXIMATE). 1338
 CC 163 (APPROXIMATE). 1346
 CC 164 (APPROXIMATE). 1354
 CC 165 (APPROXIMATE). 1362
 CC 166 (APPROXIMATE). 1370
 CC 167 (APPROXIMATE). 1378
 CC 168 (APPROXIMATE). 1386
 CC 169 (APPROXIMATE). 1394
 CC 170 (APPROXIMATE). 1402
 CC 171 (APPROXIMATE). 1410
 CC 172 (APPROXIMATE). 1418
 CC 173 (APPROXIMATE). 1426
 CC 174 (APPROXIMATE). 1434
 CC 175 (APPROXIMATE). 1442
 CC 176 (APPROXIMATE). 1450
 CC 177 (APPROXIMATE). 1458
 CC 178 (APPROXIMATE). 1466
 CC 179 (APPROXIMATE). 1474
 CC 180 (APPROXIMATE). 1482
 CC 181 (APPROXIMATE). 1490
 CC 182 (APPROXIMATE). 1498
 CC 183 (APPROXIMATE). 1506
 CC 184 (APPROXIMATE). 1514
 CC 185 (APPROXIMATE). 1522
 CC 186 (APPROXIMATE). 1530
 CC 187 (APPROXIMATE). 1538
 CC 188 (APPROXIMATE). 1546
 CC 189 (APPROXIMATE). 1554
 CC 190 (APPROXIMATE). 1562
 CC 191 (APPROXIMATE). 1570
 CC 192 (APPROXIMATE). 1578
 CC 193 (APPROXIMATE). 1586
 CC 194 (APPROXIMATE). 1594
 CC 195 (APPROXIMATE). 1602
 CC 196 (APPROXIMATE). 1610
 CC 197 (APPROXIMATE). 1618
 CC 198 (APPROXIMATE). 1626
 CC 199 (APPROXIMATE). 1634
 CC 200 (APPROXIMATE). 1642
 CC 201 (APPROXIMATE). 1650
 CC 202 (APPROXIMATE). 1658
 CC 203 (APPROXIMATE). 1666
 CC 204 (APPROXIMATE). 1674
 CC 205 (APPROXIMATE). 1682
 CC 206 (APPROXIMATE). 1690
 CC 207 (APPROXIMATE). 1698
 CC 208 (APPROXIMATE). 1706
 CC 209 (APPROXIMATE). 1714
 CC 210 (APPROXIMATE). 1722
 CC 211 (APPROXIMATE). 1730
 CC 212 (APPROXIMATE). 1738
 CC 213 (APPROXIMATE). 1746
 CC 214 (APPROXIMATE). 1754
 CC 215 (APPROXIMATE). 1762
 CC 216 (APPROXIMATE). 1770
 CC 217 (APPROXIMATE). 1778
 CC 218 (APPROXIMATE). 1786
 CC 219 (APPROXIMATE). 1794
 CC 220 (APPROXIMATE). 1802
 CC 221 (APPROXIMATE). 1810
 CC 222 (APPROXIMATE). 1818
 CC 223 (APPROXIMATE). 1826
 CC 224 (APPROXIMATE). 1834
 CC 225 (APPROXIMATE). 1842
 CC 226 (APPROXIMATE). 1850
 CC 227 (APPROXIMATE). 1858
 CC 228 (APPROXIMATE). 1866
 CC 229 (APPROXIMATE). 1874
 CC 230 (APPROXIMATE). 1882
 CC 231 (APPROXIMATE). 1890
 CC 232 (APPROXIMATE). 1898
 CC 233 (APPROXIMATE). 1906
 CC 234 (APPROXIMATE). 1914
 CC 235 (APPROXIMATE). 1922
 CC 236 (APPROXIMATE). 1930
 CC 237 (APPROXIMATE). 1938
 CC 238 (APPROXIMATE). 1946
 CC 239 (APPROXIMATE). 1954
 CC 240 (APPROXIMATE). 1962
 CC 241 (APPROXIMATE). 1970
 CC 242 (APPROXIMATE). 1978
 CC 243 (APPROXIMATE). 1986
 CC 244 (APPROXIMATE). 1994
 CC 245 (APPROXIMATE). 2002
 CC 246 (APPROXIMATE). 2010
 CC 247 (APPROXIMATE). 2018
 CC 248 (APPROXIMATE). 2026
 CC 249 (APPROXIMATE). 2034
 CC 250 (APPROXIMATE). 2042
 CC 251 (APPROXIMATE). 2050
 CC 252 (APPROXIMATE). 2058
 CC 253 (APPROXIMATE). 2066
 CC 254 (APPROXIMATE). 2074
 CC 255 (APPROXIMATE). 2082
 CC 256 (APPROXIMATE). 2090
 CC 257 (APPROXIMATE). 2098
 CC 258 (APPROXIMATE). 2106
 CC 259 (APPROXIMATE). 2114
 CC 260 (APPROXIMATE). 2122
 CC 261 (APPROXIMATE). 2130
 CC 262 (APPROXIMATE). 2138
 CC 263 (APPROXIMATE). 2146
 CC 264 (APPROXIMATE). 2154
 CC 265 (APPROXIMATE). 2162
 CC 266 (APPROXIMATE). 2170
 CC 267 (APPROXIMATE). 2178
 CC 268 (APPROXIMATE). 2186
 CC 269 (APPROXIMATE). 2194
 CC 270 (APPROXIMATE). 2202
 CC 271 (APPROXIMATE). 2210
 CC 272 (APPROXIMATE). 2218
 CC 273 (APPROXIMATE). 2226
 CC 274 (APPROXIMATE). 2234
 CC 275 (APPROXIMATE). 2242
 CC 276 (APPROXIMATE). 2250
 CC 277 (APPROXIMATE). 2258
 CC 278 (APPROXIMATE). 2266
 CC 279 (APPROXIMATE). 2274
 CC 280 (APPROXIMATE). 2282
 CC 281 (APPROXIMATE). 2290
 CC 282 (APPROXIMATE). 2298
 CC 283 (APPROXIMATE). 2306
 CC 284 (APPROXIMATE). 2314
 CC 285 (APPROXIMATE). 2322
 CC 286 (APPROXIMATE). 2330
 CC 287 (APPROXIMATE). 2338
 CC 288 (APPROXIMATE). 2346
 CC 289 (APPROXIMATE). 2354
 CC 290 (APPROXIMATE). 2362
 CC 291 (APPROXIMATE). 2370
 CC 292 (APPROXIMATE). 2378
 CC 293 (APPROXIMATE). 2386
 CC 294 (APPROXIMATE). 2394
 CC 295 (APPROXIMATE). 2402
 CC 296 (APPROXIMATE). 2410
 CC 297 (APPROXIMATE). 2418
 CC 298 (APPROXIMATE). 2426
 CC 299 (APPROXIMATE). 2434
 CC 300 (APPROXIMATE). 2442
 CC 301 (APPROXIMATE). 2450
 CC 302 (APPROXIMATE). 2458
 CC 303 (APPROXIMATE). 2466
 CC 304 (APPROXIMATE). 2474
 CC 305 (APPROXIMATE). 2482
 CC 306 (APPROXIMATE). 2490
 CC 307 (APPROXIMATE). 2498
 CC 308 (APPROXIMATE). 2506
 CC 309 (APPROXIMATE). 2514
 CC 310 (APPROXIMATE). 2522
 CC 311 (APPROXIMATE). 2530
 CC 312 (APPROXIMATE). 2538
 CC 313 (APPROXIMATE). 2546
 CC 314 (APPROXIMATE). 2554
 CC 315 (APPROXIMATE). 2562
 CC 316 (APPROXIMATE). 2570
 CC 317 (APPROXIMATE). 2578
 CC 318 (APPROXIMATE). 2586
 CC 319 (APPROXIMATE). 2594
 CC 320 (APPROXIMATE). 2602
 CC 321 (APPROXIMATE). 2610
 CC 322 (APPROXIMATE). 2618
 CC 323 (APPROXIMATE). 2626
 CC 324 (APPROXIMATE). 2634
 CC 325 (APPROXIMATE). 2642
 CC 326 (APPROXIMATE). 2650
 CC 327 (APPROXIMATE). 2658
 CC 328 (APPROXIMATE). 2666
 CC 329 (APPROXIMATE). 2674
 CC 330 (APPROXIMATE). 2682
 CC 331 (APPROXIMATE). 2690
 CC 332 (APPROXIMATE). 2698
 CC 333 (APPROXIMATE). 2706
 CC 334 (APPROXIMATE). 2714
 CC 335 (APPROXIMATE). 2722
 CC 336 (APPROXIMATE). 2730
 CC 337 (APPROXIMATE). 2738
 CC 338 (APPROXIMATE). 2746
 CC 339 (APPROXIMATE). 2754
 CC 340 (APPROXIMATE). 2762
 CC 341 (APPROXIMATE). 2770
 CC 342 (APPROXIMATE). 2778
 CC 343 (APPROXIMATE). 2786
 CC 344 (APPROXIMATE). 2794
 CC 345 (APPROXIMATE). 2802
 CC 346 (APPROXIMATE). 2810
 CC 347 (APPROXIMATE). 2818
 CC 348 (APPROXIMATE). 2826
 CC 349 (APPROXIMATE). 2834
 CC 350 (APPROXIMATE). 2842
 CC 351 (APPROXIMATE). 2850
 CC 352 (APPROXIMATE). 2858
 CC 353 (APPROXIMATE). 2866
 CC 354 (APPROXIMATE). 2874
 CC 355 (APPROXIMATE). 2882
 CC 356 (APPROXIMATE). 2890
 CC 357 (APPROXIMATE). 2898
 CC 358 (APPROXIMATE). 2906
 CC 359 (APPROXIMATE). 2914
 CC 360 (APPROXIMATE). 2922
 CC 361 (APPROXIMATE). 2930
 CC 362 (APPROXIMATE). 2938
 CC 363 (APPROXIMATE). 2946
 CC 364 (APPROXIMATE). 2954
 CC 365 (APPROXIMATE). 2962
 CC 366 (APPROXIMATE). 2970
 CC 367 (APPROXIMATE). 2978
 CC 368 (APPROXIMATE). 2986
 CC 369 (APPROXIMATE). 2994
 CC 370 (APPROXIMATE). 3002
 CC 371 (APPROXIMATE). 3010
 CC 372 (APPROXIMATE). 3018
 CC 373 (APPROXIMATE). 3026
 CC 374 (APPROXIMATE). 3034
 CC 375 (APPROXIMATE). 3042
 CC 376 (APPROXIMATE). 3050
 CC 377 (APPROXIMATE). 3058
 CC 378 (APPROXIMATE). 3066
 CC 379 (APPROXIMATE). 3074
 CC 380 (APPROXIMATE). 3082
 CC 381 (APPROXIMATE). 3090
 CC 382 (APPROXIMATE). 3098
 CC 383 (APPROXIMATE). 3106
 CC 384 (APPROXIMATE). 3114
 CC 385 (APPROXIMATE). 3122
 CC 386 (APPROXIMATE). 3130
 CC 387 (APPROXIMATE). 3138
 CC 388 (APPROXIMATE). 3146
 CC 389 (APPROXIMATE). 3154
 CC 390 (APPROXIMATE). 3162
 CC 391 (APPROXIMATE). 3170
 CC 392 (APPROXIMATE). 3178
 CC 393 (APPROXIMATE). 3186
 CC 394 (APPROXIMATE). 3194
 CC 395 (APPROXIMATE). 3202
 CC 396 (APPROXIMATE). 3210
 CC 397 (APPROXIMATE). 3218
 CC 398 (APPROXIMATE). 3226
 CC 399 (APPROXIMATE). 3234
 CC 400 (APPROXIMATE). 3242
 CC 401 (APPROXIMATE). 3250
 CC 402 (APPROXIMATE). 3258
 CC 403 (APPROXIMATE). 3266
 CC 404 (APPROXIMATE). 3274
 CC 405 (APPROXIMATE). 3282
 CC 406 (APPROXIMATE). 3290
 CC 407 (APPROXIMATE). 3298
 CC 408 (APPROXIMATE). 3306
 CC 409 (APPROXIMATE). 3314
 CC 410 (APPROXIMATE). 3322
 CC 411 (APPROXIMATE). 3330
 CC 412 (APPROXIMATE). 3338
 CC 413 (APPROXIMATE). 3346
 CC 414 (APPROXIMATE). 3354
 CC 415 (APPROXIMATE). 3362
 CC 416 (APPROXIMATE). 3370
 CC 417 (APPROXIMATE). 3378
 CC 418 (APPROXIMATE). 3386
 CC 419 (APPROXIMATE). 3394
 CC 420 (APPROXIMATE). 3402
 CC 421 (APPROXIMATE). 3410
 CC 422 (APPROXIMATE). 3418
 CC 423 (APPROXIMATE). 3426
 CC 424 (APPROXIMATE). 3434
 CC 425 (APPROXIMATE). 3442
 CC 426 (APPROXIMATE). 3450
 CC 427 (APPROXIMATE). 3458
 CC 428 (APPROXIMATE). 3466
 CC 429 (APPROXIMATE). 3474
 CC 430 (APPROXIMATE). 3482
 CC 431 (APPROXIMATE). 3490
 CC 432 (APPROXIMATE). 3498
 CC 433 (APPROXIMATE). 3506
 CC 434 (APPROXIMATE). 3514
 CC 435 (APPROXIMATE). 3522
 CC 436 (APPROXIMATE). 3530
 CC 437 (APPROXIMATE). 3538
 CC 438 (APPROXIMATE). 3546
 CC 439 (APPROXIMATE). 3554
 CC 440 (APPROXIMATE). 3562
 CC 441 (APPROXIMATE). 3570
 CC 442 (APPROXIMATE). 3578
 CC 443 (APPROXIMATE). 3586
 CC 444 (APPROXIMATE). 3594
 CC 445 (APPROXIMATE). 3602
 CC 446 (APPROXIMATE). 3610
 CC 447 (APPROXIMATE). 3618
 CC 448 (APPROXIMATE). 3626
 CC 449 (APPROXIMATE). 3634
 CC 450 (APPROXIMATE). 3642
 CC 451 (APPROXIMATE). 3650
 CC 452 (APPROXIMATE). 3658
 CC 453 (APPROXIMATE). 3666
 CC 454 (APPROXIMATE). 3674
 CC 455 (APPROXIMATE). 3682
 CC 456 (APPROXIMATE). 3690
 CC 457 (APPROXIMATE). 3698
 CC 458 (APPROXIMATE). 3706
 CC 459 (APPROXIMATE). 3714
 CC 460 (APPROXIMATE). 3722
 CC 461 (APPROXIMATE). 3730
 CC 462 (APPROXIMATE). 3738
 CC 463 (APPROXIMATE). 3746
 CC 464 (APPROXIMATE). 3754
 CC 465 (APPROXIMATE). 3762
 CC 466 (APPROXIMATE). 3770
 CC 467 (APPROXIMATE). 3778
 CC 468 (APPROXIMATE). 3786
 CC 469 (APPROXIMATE). 3794
 CC 470 (APPROXIMATE). 3802
 CC 471 (APPROXIMATE). 3810
 CC 472 (APPROXIMATE). 3818
 CC 473 (APPROXIMATE). 3826
 CC 474 (APPROXIMATE). 3834
 CC 475 (APPROXIMATE). 3842
 CC 476 (APPROXIMATE). 3850
 CC 477 (APPROXIMATE). 3858
 CC 478 (APPROXIMATE). 3866
 CC 479 (APPROXIMATE). 3874
 CC 480 (APPROXIMATE). 3882
 CC 481 (APPROXIMATE). 3890
 CC 482 (APPROXIMATE). 3898
 CC 483 (APPROXIMATE). 3906
 CC 484 (APPROXIMATE). 3914
 CC 485 (APPROXIMATE). 3922
 CC 486 (APPROXIMATE). 3930
 CC 487 (APPROXIMATE). 3938
 CC 488 (APPROXIMATE). 3946
 CC 489 (APPROXIMATE). 3954
 CC 490 (APPROXIMATE). 3962
 CC 4

FT STRAND 184 186
FT TURN 193 194
FT STRAND 196 203
FT STRAND 207 212
FT TURN 213 214
FT STRAND 215 221
FT HELIX 227 229
FT STRAND 232 237
FT STRAND 239 249
SQ SEQUENCE 249 AA; 26057 MW; ACBS42752D137650 CRC64;

Query Match 96.5%; Score 1310; DB 1; Length 249;
Best Local Similarity 98.4%; Pred. No. 1.2e-74;
Matches 245; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 2 ADNFSILHDLSSGSGNPNQGMFGAWGNOPAGAGYPGASYP-YGQAPPGAYPGQAPPG 60
DB 1 ADNFSILHDLSSGSGNPNQGMFGAWGNOPAGAGYPGASYPGAYPGQAPPGAYPGQAPPG 60
QY 61 AVHGAPGAYPGAPGAYPGGPGSGGAYPGSSGQSPAPGAY-ATGPGAPAGPLIYVNL 119
DB 61 AVHGAPGAYPGAPGAYPGGPGSGGAYPGSSGQSPAPGAYATGPGAPAGPLIYVNL 120
QY 120 LPGVVPRMLITILIGTVKPNANRIALDFQRGNDVAFHF-PRPNENNRRTVTCNTKLDNNW 178
DB 121 LPGVVPRMLITILIGTVKPNANRIALDFQRGNDVAFHFPRPNENNRRTVTCNTKLDNNW 180
QY 179 GREEROSVPPESGKPFKIOVLVEPDHFVAVNDAAH-LOYNHRVKLNLSKLGISGDI 237
DB 181 GREEROSVPPESGKPFKIOVLVEPDHFVAVNDAAHLOYNHRVKLNLSKLGISGDI 240
QY 238 LTSASYTMI 246
DB 241 LTSASYTMI 249

RESULT 2
LEG3 CANFA STANDARD; PRT; 295 AA.

AC P38486:
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Galectin-3 (Galectose-specific lectin 3) (MAC-2 antigen) (IGF-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35)
DE (Laminin-binding protein) (lectin L-29).
GN LGALS3.

OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxID=9615;
RN [1]
RN SEQUENCE OF 11-295 FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Kidney epithelium;
RX MEDLINE=94075368; PubMed=8253805;
RX Hermann U., Turck C.W., Atchison R.E., Huflejt M.E., Poulter L.,
RA Gilt M.A., Burlingame A.L., Barondes S.H., Loeffler H.;
RT "Primary structure of the soluble lactose binding lectin L-29 from
rat and dog and interaction of its non-collagenous proline-,
glycine-, tyrosine-rich sequence with bacterial and tissue
collagenase.";
RT J. Biol. Chem. 268:26704-26711(1993).

RL [2]
RP PHOSPHORYLATION.

RX MEDLINE=94075369; PubMed=8253806;
RX Huflejt M.E., Turck C.W., Lindsed R., Barondes S.H., Loeffler H.;
RT "L-29, a soluble lactose-binding lectin, is phosphorylated on serine
6 and serine 12 in vivo and by casein kinase I.";
RL J. Biol. Chem. 268:26712-26718(1993).

CC -1- FUNCTION: Galactose-specific lectin which binds Igf.
CC -1- SUBUNIT: Probably forms homo- or heterodimers.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. May be secreted by
a non-classical secretory pathway.

CC -1- PTM: The degree of phosphorylation is higher in the cytoplasmic
form than in the nuclear form.
CC -1- SIMILARITY: In the C-terminal section; belongs to the galectin
CC (galactin/S-lectin) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
or send an email to license@isb-sdb.ch).

CC EMBL; L23429; AAA16211.1; -
DR HSSP; E17931; 1A3K.
DR InterPro; IPR009895; Cons. like _lec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
KW Galectin; lectin; IGF-binding protein; Repeat; Phosphorylation;
KW Acetylation; Nuclear protein.

FT INTL_MET 0 0
FT MOD_RES 1 1 ACETYLATION (BY CK1) (90%).
FT MOD_RES 5 5 PHOSPHORYLATION (BY CK1) (10%).
FT MOD_RES 11 11 12 X 9 AA TANDEM REPEATS OF Y-P-G-X(3) -P-
FT DOMAIN 35 142 G-[GAT].

FT REPEAT 35 43 1.
FT REPEAT 44 52 2.
FT REPEAT 53 61 3.
FT REPEAT 62 70 4.
FT REPEAT 71 79 5.
FT REPEAT 80 88 6.
FT REPEAT 89 97 7.
FT REPEAT 98 106 8. (APPROXIMATE).
FT REPEAT 107 114 9. (APPROXIMATE).
FT REPEAT 115 123 10.
FT REPEAT 124 133 11 (APPROXIMATE).
FT REPEAT 134 142 12 (APPROXIMATE).
FT DOMAIN 152 295 GALAPTIN.
FT DISULFID 218 218 INTERCHAIN (BY SIMILARITY).
FT BINDING 226 232 BRTA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 295 AA; 30199 MW; 4921327063CE41C7 CRC64;

Query Match 80.4%; Score 1091; DB 1; Length 295;
Best Local Similarity 71.2%; Pred. No. 4.6e-61;
Matches 210; Conservative 13; Mismatches 22; Indels 50; Gaps 6;

QY 2 ADNFSILHDLSSGSGNPNQGMFGAWGNOPAGAGYPGASYP----- 42
DB 1 ADNFSILHDLSSGSGNPNQGMFGAWGNOPAGAGYPGASYPGAYPGQAPPGAYPGQAPPG 60
QY 43 -----GYGQAPPGAYPGQAPPGAYHG--AGAYPGAPGAYPP----- 81
DB 61 GYPGQAPPGYPGQAPPGYPGQAPPGYPGQAPPGYPGQAPPGYPGQAPPGYPGQAPPGY 120
QY 82 -----PSGPGAYSSGQSPAPGAY-ATGPGAPAGPLIYVNLPLPGVPRMLITIL 133
DB 121 PGTOGQPSGPGAYPPPGQSPAPGAYPAAGPGIAGPLIYVNDPLPGVPRMLITIL 180
QY 134 GTVKENANRIALDFQRGNDVAFHF-PRPNENNRRTVTCNTKLDNNWGREEROSVPPESG 192
DB 181 GTVRSANRIALDFQRGNDVAFHFPRPNEDNRRTVTCNTKLDNNWGREERQAPFESG 240
QY 193 KFFKIQVLVEPDHFVAVNDAAH-LOYNHRVKLNLSKLGISGDI DTLSASYTMI 246
DB 241 KFFKIQVLVESDHFVAVNDAAHLOYNHRVKLNLSKLGISGDI DTLSASYTMI 295

RESULT 3
LEG3_MOUSE STANDARD; PRT; 263 AA.

AC P16110; 01-APR-1990 (Rel. 14, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Galactin-3 (galactose-specific lectin 3) (Mac-2 antigen) (IgF-binding protein) (35 kDa lectin) (carbohydrate binding protein 35) (CBP 35) (laminin-binding protein) (lectin L-29) (L-34 galactoside-binding lectin).
 DE LgalS3.
 GN LgalS3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBA/2J; TISSUE=Macrophage;
 RX MEDLINE=90063462; PubMed=2584931;
 RA Cherayil B.U., Weiner S.U., Pilla S.;
 RT "The Mac-2 antigen is a galactose-specific lectin that binds IgF";
 RL J. Exp. Med. 170:11959-1972(1989).
 [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=68198129; PubMed=3360772;
 RA Jia S., Wang J.L.;
 RT "Carbohydrate binding protein 35. Complementary DNA sequence reveals homology with proteins of the heterogeneous nuclear RNP";
 RL J. Biol. Chem. 263:6009-6011(1988).
 [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=69275058; PubMed=2525069;
 RA Raz A., Pazerini G., Carmi P.;
 RT "Identification of the metastasis-associated, galactoside-binding lectin as a chimeric gene product with homology to an IgF-binding protein";
 RL Cancer Res. 49:3489-3493(1989).
 [4]
 RP SEQUENCE OF 158-162; 165-174 AND 213-225.
 RC MEDLINE=90236991; PubMed=2332426;
 RA Woo H.-U., Shaw J.M., Messier J.M., Mercurio A.M.;
 RT "The major non-integer laminin binding protein of macrophages is identical to carbohydrate binding protein 35 (Mac-2).";
 RL J. Biol. Chem. 265:7097-7099(1990).
 [5]
 RP DISULFIDE BOND.
 RC MEDLINE=92011585; PubMed=1917966;
 RA Woo H.-U., Lotz M.M., Jung J.U., Mercurio A.M.;
 RT "Carbohydrate-binding protein 35 (Mac-2), a laminin-binding lectin, forms functional dimers using cysteine 186.";
 RL J. Biol. Chem. 266:18419-18422(1991).
 CC -1- FUNCTION: Galactose-specific lectin which binds IgE.
 CC -1- SUBUNIT: Exists as several distinct species of 35, 67 and 80 kDa. Probably forms homo- or heterodimers.
 CC -1- TISSUE SPECIFICITY: The highest levels are found in activated macrophages.
 CC -1- SIMILARITY: In the C-terminal section; belongs to the galectin (galactin/S-lectin) family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/annouce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X15834; CAA34736.1; -
 DR EMBL; U03723; AAA37311.1; -
 DR EMBL; X16074; CAA34206.1; -
 DR PIR; S08537; A28651.
 DR HSSP; P17931; 1A3K.
 DR MGD; MGI:96778; LgalS3.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.

Query	Match	Score	DB 1	Length	263
Query	Best Local Similarity	79.4%			
Matches	208	Conservative	16	Mismatches	17
				Indels	26
				Gaps	8
QY	2	ADNFSLDALSGSNPNPCGMPGAGGPGAGSYPG-YPGQAPGAYPGQAPPG	60		
DB	1	ADGSLINDALGSGNPNPQGGPGAGNQP-GAGGPGAGAYPGAGQAPGAYPGQAPPG	59		
QY	61	AYHG-----APGAYPGAPGYPGPPSGPGAYSSGGQPSAPGAY-----A	101		
DB	60	AYPGQAPPSAYPGPTAGAYGPTAGAYPGQQA-PGAF-P--GGPAGAPGQSGGYPA	116		
QY	102	TGPFGAGAPGLIVYNNPLPFGVPRMLITLIGYKXNARNIADPQAGDVAFHF-PRF	160		
DB	117	AGPGVAGPLVAYDPLPGVWRMLITMGVKXNARIVDFRGNDAVAFHMERF	176		
QY	161	NENNRATVYCNCTKLDNNMGREERQSVPPESGKRFKIQVLEPHEFKVAVNDH-LOYNH	219		
DB	177	NENNRATVYCNCTKDDNNMGKEERQSAFPFBSGKRFKIQVLEALHFKVAVNDHLLQYNH	236		
QY	220	RVKKLNFIKSLGISGIDLTLSASYTMT	246		
DB	237	RMKRLREISQLGISGIDLTLSANHAMT	263		
RESULT 4					
LEG3	RAT	STANDARD	PRT	261	AA.
AC	P06699				
DT	01-JAN-1988	(Rel. 06, Created)			
DT	01-NOV-1995	(Rel. 32, last sequence update)			
DT	15-MAR-2004	(Rel. 43, last annotation update)			
DE	Galectin-3 (galactose-specific lectin 3) (MAC-2 antigen) (IGF-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35)				
DE	(laminin-binding protein) (Lectin I-29).				
GN	IGALB3.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
NCBI	NCBI_Taxid=10116;				
LN	[1]				

RP SEQUENCE FROM N.A.
 RX MEDLINE=88016189; PubMed=2958848;
 RA Albrandt K., Orida N.K., Liu F.-T.;
 RT "An Ige-binding protein with a distinctive repetitive sequence and
 RT homology with an Igc receptor."; Proc. Natl. Acad. Sci. U.S.A. 84:6859-6863(1987).
 RN [2]
 RN SEQUENCE OF 124-261 FROM N.A.
 RX MEDLINE=85216641; PubMed=3858867;
 RA Liu F.-T., Albrandt K., Mendel E., Kulczycki A. Jr., Orida N.K.;
 RT "Identification of an Ige-binding protein by molecular cloning."; Proc. Natl. Acad. Sci. U.S.A. 82:4100-4104(1985).
 RN [3]
 RP SEQUENCE OF 119-144.
 RX MEDLINE=90105471; PubMed=2605254;
 RA Lefler H., Masiarz P.R., Barondes S.H.;
 RT "Soluble lactose-binding vertebrate lectins: a growing family."; Biochemistry 28:9222-9229(1989).
 RN [4]
 RP PARTIAL SEQUENCE, AND ACETYLATION.
 RX MEDLINE=94075368; PubMed=8253805;
 RA Hermann J., Turck C.W., Archison R.E., Huflejt M.E., Poulter L., Gilt M.A., Burlingame A.L., Barondes S.H., Lefler H.;
 RT "Primary structure of the soluble lactose binding lectin L-29 from rat and dog and interaction of its non-collagenous proline-, glycine-, tyrosine-rich sequence with bacterial and tissue collagenase."; J. Biol. Chem. 268:26704-26711(1993).
 CC -1- SUBUNIT: Galactose-specific lectin which binds IGE.
 CC -1- SIMILARITY: Probably forms homo- or heterodimers.
 CC -1- SIMILARITY: In the C-terminal section; belongs to the galectin (galactin/S-lectin) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, J02962; AAA40828.1; -;
 DR EMBL, M13697; AAA41378.1; -;
 DR PIR, A54889; A54889.
 DR HSSP, P17931; 1A3X.
 DR InterPro; IPR008985; Cona_1like_lect_gl.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 1.
 DR PROSITE; PS00309; GALAPTIN; 1.
 DR KX Galectin; Lectin; IGE-binding protein; Repeat; Phosphorylation;
 KW Acetylation.
 FT INIT_MET 0 0
 FT MOD_RES 1 1
 FT MOD_RES 5 5
 FT DOMAIN 34 111
 FT REPEAT 34 42
 FT REPEAT 43 51
 FT REPEAT 52 60
 FT REPEAT 61 69
 FT REPEAT 70 78
 FT REPEAT 79 87
 FT REPEAT 88 97
 FT REPEAT 98 104
 FT REPEAT 105 111
 FT DOMAIN 129 261
 FT DISULFID 184 184
 FT BONDING 192 198
 FT CONFLICT 19 19
 FT SEQUENCE 261 AA, 27070 KW, EAPFAL17F5EA5080D CRC64;

Query Match 79.1%; Score 1074; DB 1; Length 261;
 Best Local Similarity 77.9%; Pred. No. 4.6e-60;
 Matches 204; Conservative 16; Mismatches 24; Indels 18; Gaps 7;
 QY 2 ADNSLHDALSGSNPNQCPGAWGNOPAGAGYPCASITG-YPGCAPPGAYRQAPRG 60
 DB 1 ADGFSLNDALAGSNPNQCPGAWGNOP-GAGGYPCASITGAYPGQAPPGYRQAPPS 59
 QY 61 AYHG-----APGAYGAPAGYVPGSPGPGAYPSS--GQSPAGAY-ATGPGY 106
 DB 60 AYPGPTGSAVPGPTAPGAFPGQPGGAFPSAPGAYPSAPGAYPATGPG 119
 QY 107 APAGLIVPNLPLPGGVPPMLITITGVKPNNRIALDFQGNDAVHR-PRNENNR 165
 DB 120 APTGGLIVPNMPLPGGVPPMLITITGVKPNNSITLNKKGNDIAFHPRNENNR 179
 QY 166 RVIYCNITKLDNNKGREERQSVPPESGKPEKIQVIVEPDHKKVAVNDNH-LOYNHRVKKL 224
 DB 180 RVIYCNTRQDNNMGREERQSAFPESGKPEKIQVIVEADHKKVAVNDVHLIQYHNRKNTL 239
 QY 225 NEISKLTGSGDIDLTSAATMI 246
 DB 240 REISQLGITITLTSASHMI 261
 RESULT 5
 LEG3 CRITO STANDARD; PRT; 244 AA.
 AC P47953;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Galectin-3 (Galactose-specific lectin 3) (MNC-2 antigen) (IGE-binding protein) (35 kDa lectin) (carbohydrate binding protein 35) (CBP 35) (Laminin-binding protein) (Lectin L-29) (CBP30).
 GN LGALS3.
 OS Cricetus longicaudatus (long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Cricetus.
 OX NCBI_TaxID=10030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=94293546; PubMed=8027086;
 RA Mehrl B., Bayumia S., Martin S.R., Hughes R.C.;
 RT "Structure of baby hamster kidney carbohydrate-binding protein CBP30, an S-type animal lectin."; J. Biol. Chem. 269:18250-18258(1994).
 RL -1- FUNCTION: Galactose-specific lectin which binds IGE.
 CC -1- SIMILARITY: In the C-terminal section; belongs to the galectin (galactin/S-lectin) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X78879; CAA55479.1; -;
 DR HSSP, P17931; 1A3X.
 DR InterPro; IPR008985; Cona_1like_lect_gl.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 1.
 DR PROSITE; PS00309; GALAPTIN; 1.
 DR KX Galectin; Lectin; IGE-binding protein; Repeat; Phosphorylation;
 KW Acetylation.
 FT INIT_MET 0 0
 FT MOD_RES 1 1
 FT MOD_RES 5 5
 FT SEQUENCE 261 AA, 27070 KW, EAPFAL17F5EA5080D CRC64;

```

      (BY SIMILARITY).
      7 X 9 AA TANDEM REPEATS OF Y-P-G-X(3)-P-
      [GS]-A.
      FT DOMAIN 34 98
      FT REPEAT 34 42
      FT REPEAT 43 51
      FT REPEAT 52 60
      FT REPEAT 61 69
      FT REPEAT 70 77
      FT REPEAT 78 87
      FT REPEAT 88 98
      FT DOMAIN 112 244
      FT DISULFID 167 181
      FT BINDING 175 181
      SQ SEQUENCE 244 AA: 25608 MW, 8F99B9AA0BBA7D3F CRC64;

Query Match 77.0%; Score 1044.5; DB 1; Length 244;
Best Local Similarity 80.9%; Pred. No. 2.9e-58;
Matches 203; Conservative 12; Mismatches 23; Indels 13; Gaps 8;

QY 2 ADNFSLHDALSSGNPNPGMNGNPGAGYPGASYPG-YPGQAPPGAPPGAPG 60
DB 1 ADGFSINDALSSGNPNPGMNGNPGAGYPGASYPG-YPGQAPPGAPPGAPG 59
QY 61 AYHG-APGAYPGAPGAYPGGPGGAYPSSGQSPAGAY-ATGPGAPAGPLIYPN 117
DB 60 AYPGTPAPGAYPG-PAPGAYPGQPGASGAY-----PSAGAYPAGPYGAPGALTVPYK 113
QY 118 LPLPGSVPRMLITIGTKENANRALDPRGNDVAHFH-PRFENNRRVYICNTKLDN 176
DB 114 LPLAGVPMRMLITIGTKENANRALDPRGNDVAHFHPRFENNRRVYICNTKLDN 173
QY 177 NMGRERQSVPPESGPKFQIVLVEPDHFKVAVNDAH-LQYNHRVKLNISKLGISGD 235
DB 174 NMGRERQSAFPESGPKFQIVLVEADHFKVAVNDAHLLQYNHRMKNLREINQMEISGD 233
QY 236 IDLTASATYMI 246
DB 234 ILTLSAATMI 244

RESULT 6
LEG3 RABIT STANDARD; PRT; 241 AA.
ID LEG3 RABIT STANDARD; PRT; 241 AA.
AC P47845;
DB 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-3 (Galactose-specific lectin 3) (Mac-2 antigen) (IGE-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35)
DE (Laminin-binding protein) (Lectin L-29).
DE IGALS3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Aorta;
RX MEDLINE=96011642; PubMed=7590275;
RA Gaudin J.-C., Monsigny M., Legendre A.,
RT "Cloning of the cDNA encoding rabbit galectin-3."
RL Gene 163:249-252(1995).
CC -1- FUNCTION: Galactose-specific lectin which binds IGE.
CC -1- SIMILARITY: In the C-terminal section, belongs to the galectin (galactin/5-lectin) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

```

```

      EMBL; U06470; AAC8491.1; -.
      DR PIR; J04300; J04300.
      DR HSSP; P17931; IABK.
      DR InterPro; IPR008985; Cona1_1ec_g1.
      DR InterPro; IPR01079; Galectin.
      DR Pfam; PF00337; Gal-bind_lectin; 1.
      DR SMART; SM00276; GLECT_1.
      DR PROSITE; PS00309; Galactin; 1.
      DR Galectin; Lectin; IGE-binding protein; Repeat; Phosphorylation;
      KW Acetylation.
      FT INIT MET 0 0
      FT MOD_RES 1 1
      FT MOD_RES 5 5
      FT MOD_RES 5 5
      FT DOMAIN 34 97
      FT REPEAT 34 42
      FT REPEAT 43 51
      FT REPEAT 52 60
      FT REPEAT 61 69
      FT REPEAT 70 79
      FT REPEAT 80 91
      FT REPEAT 92 97
      FT DOMAIN 109 241
      FT DISULFID 164 164
      FT BINDING 172 180
      SQ SEQUENCE 241 AA: 25370 MW, 6EFC42D626479F9 CRC64;

Query Match 75.1%; Score 1019; DB 1; Length 241;
Best Local Similarity 78.7%; Pred. No. 1.1e-56;
Matches 196; Conservative 15; Mismatches 26; Indels 12; Gaps 7;

QY 2 ADNFSLHDALSSGNPNPGMNGNPGAGYPGASYPG-YPGQAPPGAPPGAPG 60
DB 1 ADGFSINDALSSGNPNPGMNGNPGAGYPGASYPG-YPGQAPPGAPPGAPG 59
QY 61 AYHGAPGAYPGAPGAYPGGPGGAYPSSGQSPAGAY-ATGPGAPAGPLIYPNLP 119
DB 60 PYPG-PGAY-----GAYPGQPGAYPSPGQSPAGAYPAGSPASAGPLPYVYDLP 112
QY 120 LPLGVPRMLITIGTKENANRALDPRGNDVAHFH-PRFENNRRVYICNTKLDNNW 178
DB 113 LPLGVPMRMLITIGTKENANRALDPRGNDVAHFHPRFENNRRVYICNTKLDNNW 172
QY 179 GREEQSVPPESGPKFQIVLVEPDHFKVAVNDAH-LQYNHRVKLNISKLGISGDID 237
DB 173 GREEQVTPFPFQIGKPKFQIVLVEPDHFKVAVNDAHLLQYNHRMKNLKEINKLGISGDIQ 232
QY 238 LTSASATYMI 246
DB 233 LTSASATMI 241

RESULT 7
LEG3 MOUSE STANDARD; PRT; 353 AA.
ID LEG3 MOUSE STANDARD; PRT; 353 AA.
AC C08573; C08572;
DB 01-NOV-1997 (Rel. 35, Created)
DB 01-NOV-1997 (Rel. 35, Last sequence update)
DB 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-9.
DE IGALS9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Schiurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney, and Small intestine;
RX MEDLINE=97190351; PubMed=9038233;
RA Wada J., Kanwar Y.S.;
RT "Identification and characterization of galectin-9, a novel beta-

```

[illegible]

```

Db      286  GQERRSLTGMPEPRGGSFSVWITCEBHCCKVAVNGMCEYXRLLKNLGDINTLEVAGD 345
OY      236  IDLT 239
          ||
Db      346  IOLT 349

RESULT 8
LEG9_HUMAN
ID      LEG9_HUMAN      STANDARD;          PRT;          355 AA.
AC      000182; 014532; 075028; GQNO58;
DT      01-NOV-1997 (Rel. 35, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Galectin-9 (HOM-4D-21) (Ecallectin).
CN      LGALS9.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM SHORT).
RC      TISSUE=Spleen;
RX      MEDLINE=97197815; Pubmed=9045665;
RA      Turecel O., Schmitt H., Padle N., Freundschuh M., Sahin U.;
RT      "Molecular definition of a novel human galectin which is immunogenic
RL      in patients with Hodgkin's disease."
RN      J. Biol. Chem. 272:6416-6422(1997).
[2]
RP      SEQUENCE FROM N.A. (ISOFORM LONG).
RC      TISSUE=Gastric carcinoma;
RA      Kato S.;
RL      Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A. (ISOFORM SHORT).
RX      MEDLINE=98307937; Pubmed=9642261;
RA      Matsumoto R., Matsumoto H., Seki M., Hata M., Asano Y., Kanegaeaki S.,
RA      Stevens R.L., Hiraishima M.;
RT      "Human ecallectin, a variant of human galectin-9, is a novel eosinophil
RL      chemoattractant produced by T lymphocytes."
RN      J. Biol. Chem. 273:16976-16984(1998).
[4]
RP      SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RA      Akiyama S.;
RT      "Homo sapiens galectin-9 (LGALS9) / ecallectin gene, exon 2 through
RL      11."
RN      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN      SEQUENCE FROM N.A. (ISOFORM LONG).
RA      Graessler J., Spitzzenberger F., Schroeder H.E.;
RT      "Genomic organization of the human galectin-9 gene";
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC      -I- FUNCTION: Birds galactosides. May play a role in thymocyte-
CC      epithelial interactions relevant to the biology of the thymus. The
CC      short isoform acts as an eosinophil chemoattractant.
CC      -I- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Comment=Additional isoforms seem to exist;
CC      Name=Long;
CC      IsoId=000182-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=000182-2; Sequence=VSP_003096;
CC      -I- TISSUE SPECIFICITY: Peripheral blood leukocytes and lymphatic
CC      tissues. Overexpressed in Hodgkin's disease tissue.
CC      -I- DOMAIN: Contains two homologous but distinct carbohydrate-binding
CC      domains.
CC      -I- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial

```

DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)

DE Galectin-9 (36 kDa beta-galactoside binding lectin) (urate
DE transporter/channel) (UAT).
CN ICALIS9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=Sprague-Dawley; TISSUE=Kidney, and Small intestine;
RX MEDLINE=97190351; Pubmed=9088233;
RA Wada J., Kanwar Y.S.;
RT "Identification and characterization of galectin-9, a novel beta-
RT galactoside-binding mammalian lectin.";
RL J. Biol. Chem. 272:6078-6086(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=Sprague-Dawley; TISSUE=Kidney;
CC

```

RX MEDLINE=97150769; Pubmed=99963305;
RA Igal-Pinto E., Tao W., Rappaport U., Richardson M., Knorr B.A.,
RT Abramson R.G.;
RT "Molecular cloning and functional reconstitution of a urate
RT transporter/channel.";
RL J. Biol. Chem. 272:617-625 (1997).
CC -1- FUNCTION: Binds galactosides. May play a role in thymocyte-
CC epithelial interactions relevant to the biology of the thymus (By
CC similarity). May provide the molecular basis for urate flux across
CC cell membranes, allowing urate that is formed during purine
CC metabolism to efflux from cells and serving as an electrogenic
CC transporter that plays an important role in renal and
CC gastrointestinal urate excretion. Highly selective to the anion
CC urate.
CC -1- SUBCELLULAR LOCATION: CYTOSOLASMIC. MAY ALSO BE SECRETED BY A NON-
CC CLASSICAL SECRETORY PATHWAY (By SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Long;
CC IsoId=P97840-1; Sequence=displayed;
CC Name=Short;
CC IsoId=P97840-2; Sequence=VSP_003098;
CC -1- TISSUE SPECIFICITY: The long form is expressed exclusively in the
CC small intestine.
CC -1- DOMAIN: Contains two homologous but distinct carbohydrate-binding

```

```

CC      domains.
CC      - - SIMILARITY: Belongs to the galectin (galptin/S-lectin) family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collabor
CC      between the Swiss Institute of Bioinformatics and the EMBL outstat
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commu
CC      entities requires a license agreement (See http://www.isb.ch/anno
CC      or send an email to license@isb-stb.ch).
CC      -----
CC      EMBL: U59462; AAB51192.1; --
CC      DR EMBL: U72741; AAB68592.1; --
CC      DR EMBL: U67958; AAB48591.1; --
CC      DR HSSP: P17931; IA3K.
CC      DR InterPro: IPR008985; Cona1_like_1ec_g1.
CC      DR InterPro: IPR001079; Galectin.
CC      DR Pfam: PF00337; Gal-binding_lectin.
CC      DR SMART: SM00276; GLECF; 2.
CC      PROSITE: PS00309; GALPTIN. 2

```

KM	Galectin; Lectin; Repeat; Alternative splicing; Ion transport.
FT	DOMAIN 1 147 GALAPTIN 1.
FT	DOMAIN 148 205 LINKER.
FT	DOMAIN 206 354 GALAPTIN 2.
FT	BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).
FT	BINDING 286 292 BETA-GALACTOSIDE (BY SIMILARITY).
FT	VARSPIC 148 179 Missing (in Isoform Short).

```

FT  SEQUENCE 354 AA; 39946 MW; 6574F960B2EAF37C CRC64; /FTID=VSP 003098.
SQ  Query Match 20.4%; Score 276.5; DB 1; Length 354;
    Best Local Similarity 37.4%; Pred. No. 1.2e-10;
    Matches 74; Conservative 29; Mismatches 74; Indels 21; Gaps 9;

QY 50 PGAVPGQAPPCAHYHGAPAYGAPAP--GVYPGPGSPGAPVPS--GQSPAPGAYATQPY 105
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 166 PVQPE-RPEKRRKQKQRTQFQALQAPVAQTIIHVHSIPGMSTPGIP--PMAPTRAY 222
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 106 GAPAGPLIVPNLPLPGVGVPRMLTIAGTPEKNARIALDFQRGNDAVAFH--PRENNN 164
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 223 -----TIPEFTSIPIPNFYFSKSNISGVLPDKKRHINMLRCGDLAFHINPFE-- 273
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 165 RRVVYCNFKLNNNGRERQ--SVPEPSGKPEFKIQLVVEPDHFKVAVNDAHL-QYNHRV 221
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 274 -KVVAVRTQIINNSWGPERSLPGRMPFRGQSFVWLICGHCFAVADGHIICHYHRL 332
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 222 KKLNEISKLGISGIDLT 239
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 333 KNLPEINTLEAVAGDIQLT 350
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
LEG4_PIG STANDARD; PRT; 323 AA.
AC Q29058; Q29296;
DT 01-NOV-1997 (Rel. 35 Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Galectin-4 (lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36LBP).
GN LGALS4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE= Tongue;
RC MEDLINE=95081129; PubMed=7989350;
RA Chiu M.L., Perry D.A.D., Feldman S.R., Klapper D.G., O'Keefe E.J.;
RT "An adherens junction protein is a member of the family of lactose-
RT binding lectins."
RT J. Biol. Chem. 269:31770-31776 (1994).
RN [2]
RP SEQUENCE OF 1-140 FROM N.A.
RC TISSUE= Small intestine;
RC MEDLINE=96327607; PubMed=8672129;
RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine CDNA
RT library: analysis of 839 clones."
RL Mamm. Genome 7:509-517 (1996).
CC -!- FUNCTION: Galectin that binds lactose and a related range of
CC sugars. May be involved in the assembly of adherens junctions.
CC -!- SUBUNIT: Monomer.
CC -!- DOMAIN: Contains two homologous but distinct carbohydrate-binding
CC domains.
CC -!- SIMILARITY: Belongs to the galectin (galaplin/S-lectin) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79303; CAA55884.1; -.
DR EMBL; F14653; CAA23179.1; -.
DR PIR; A55664; A55664.
DR HSSP; P47929; 1BKZ.

```

```

DR InterPro; IPR008985; Consilium_1ec_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; FALSE_NEG.
KW Galectin; Lectin; Repeat.
FT DOMAIN 1 152 GALAPTIN 1.
FT DOMAIN 153 177 LINER.
FT DOMAIN 178 323 GALAPTIN 2.
FT BINDING 256 262 BETA-GALACTOSIDE (BY SIMILARITY).
FT CONFLICT 98 99 AP -> GA (IN REF. 2).
FT CONFLICT 126 126 H -> T (IN REF. 2).
SQ SEQUENCE 323 AA; 35852 MW; 728C761712B29343 CRC64;

Query March 18.7%; Score 254; DB 1; Length 323;
Matches 68; Conservative 22; Mismatches 77; Indels 14; Gaps 8

QY 71 GAPAPGVYPPGPGGAPYSSGSPSAGAVATGFGYAPAGPLIVPYNLPLPGVPPRLI 130
Db 152 GQAPV--PGMPVPG--YFGGKINQPCNLCPEGAPTFPPVDPYKTRLLGGIVARRTI 208
QY 131 TIIGTVYPPNNRRLDQRCG--DYAHF--PRFENNRRVIVCTKLDNNMGRERQSVF 187
Db 209 VIKGYVPPSGKSLVINKFVSGSGVAAHINERLLEG--IVRNSYLVKGGAEERKSSF 265
QY 188 -PFESGKPFKIQVVEPDHFEEVAVNDAHL-QYNRRVKKLEISKIGSGIDLTLSASYTM 245
Db 266 NPFAPGVYFDLSTGCGIDRFKRVYVANGHLPDFSHRLNFGQVDLTLEQDPTL---SVYQ 322
QY 246 I 246
Db 323 I 323

RESULT 11
LEG3_RAT
ID LEG3_RAT STANDARD; PRT; 144 AA.
AC P47967;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Galectin-5 (RL-18).
GN LGALS5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Reticuloocytes;
RX MEDLINE=95197487; PubMed=7890611;
RA Gilt M.A., Wiser M.F., Jeffler H., Herrmann U., Xia Y.-R.,
RA Maasa S.M., Cooper D.N.W., Lusis A.J., Baronas S.H.;
RT "Sequence and mapping of galectin-5, a beta-galactoside-binding
RT lectin, found in rat erythrocytes."
RL J. Biol. Chem. 270:5032-5038(1995).
CC -1- FUNCTION: May function in erythrocyte differentiation.
CC -1- SUBUNIT: Monomer.
CC -1- TISSUE SPECIFICITY: Erythrocytes.
CC -1- SIMILARITY: Belongs to the galectin (galaptin/s-lectin) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
RR EMBL; L36862; AAC42050.1; -.
RR EMBL; L21711; AAA65445.1; -.
RR PIR; A55932; A55932.

```

```

DR HSSP; PA7929; 1BKZ.
DR InterPro; IPR008985; Cons like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
DR Galectin; Lectin; Acetylation.
FT INIT MET 0
FT MOD RES 1 1 ACETYLATION.
FT BINDING 76 82 BETA-GALACTOSIDE (POTENTIAL).
FT CONFLICT 127 135 MISSING (IN REF. 1; AA65445).
SQ SEQUENCE 144 AA; 16065 MW; BC95283D760DA515 CRC64;

Query Match 18.5%; Score 250.5; DB 1; Length 144;
Best Local Similarity 40.4%; Pred. No. 2.1e-09;
Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPNPLPGGVPRMLTITIGTVKPNANRIALDFQRGNDAVAFH-PRFNNR 166
DB 8 PYVNLAVPFTSIPNGLYPSKSIIVSGVLSDAKRFQINRCGGDIAFHLNPRFDSN---64

QY 167 VIYCNKLDNNMGRERQ--SVFPESGKPKIQLVDPHFVAVNDAL-QYNNRYKK 223
DB 65 AAVRNTQINNSWGPERSLPGSMPSRGGFVSWMICEGHCFAVADGCHI CEYSHLNM 124
QY 224 LNEISKIGISGDIIDLT 239
DB 125 LPDINTLLEVAQDIFLT 140

RESULT 12
LEG6 MOUSE STANDARD; PRT; 301 AA.
ID LEG6 MOUSE
AC 054891; O88352;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Galectin-6.
GN LGALS6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98112847; PubMed=9446608;
RA Gilt M.A., Colnot C., Poltler F., Nani K.J., Barondes S.H.,
RA Leflier H.;
RT "Galectin-4 and galectin-6 are two closely related lectins expressed
in mouse gastrointestinal tract.";
RL J. Biol. Chem. 273:2954-2960(1998).

RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98112848; PubMed=9446609;
RA Gilt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Barondes S.H.,
RA Leflier H.;
RT "Sequence, structure, and chromosomal mapping of the mouse Lgalst6
gene, encoding galectin-6.";
RL J. Biol. Chem. 273:2961-2970(1998).

-1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
domains.
-1- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

DR EMBL; AF026799; AAC04508.1; -.
DR EMBL; AF026796; AAC04508.1; JOINED.
DR EMBL; AF026797; AAC04508.1; JOINED.
DR EMBL; AF026798; AAC04508.1; JOINED.
DR EMBL; AF026794; AAC27244.1; -.
DR HSSP; P17931; LA3K.
DR MGD; MGI:107535; Lgalst6.
DR InterPro; IPR008985; Cons like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
DR Galectin; Lectin; Repeat.
FT DOMAIN 1 151 GALAPTIN 1.
FT DOMAIN 152 160 LINKER.
FT DOMAIN 160 301 GALAPTIN 2.
FT CONFLICT 154 154 A -> V (IN REF. 2; AAC27244).
SQ SEQUENCE 301 AA; 34112 MW; 9A4DD09944EDFAB9 CRC64;

Query Match 18.1%; Score 245; DB 1; Length 301;
Best Local Similarity 37.7%; Pred. No. 8.8e-09;
Matches 57; Conservative 29; Mismatches 55; Indels 10; Gaps 6;

QY 99 AYATGYGAPAGELIVPNPLPGGVPRMLTITIGTVKPNANRIALDF---QRGNDA 154
DB 2 AYVAPGYQPTYNPLTPYKRPILPGGLVGMSPFYIGTAKENMRPFVNFVAGQDDADVA 61

QY 155 FHF-PRFNNRNVYCNKLDNNMGRERQSVFPESGKPKIQLVDPHFVAVNDAL 213
DB 62 FHFNPRFDGMDR--VFENTKSGRWGKEEKSVM-PRQKGHFLVPMVEHYKVVNNS 118
QY 214 -HLYNNHRYVKLNEISKIGISGDIIDLTASAY 243
DB 119 PFVEYHRL--PYQWTHLQVDDGLDLSINF 148

RESULT 13
LEG4 RAT STANDARD; PRT; 324 AA.
ID LEG4 RAT
AC P38552;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36BP).
GN LGALS4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=93194902; PubMed=8449956;
RA Oda Y., Herrmann U., Gilt M., Turck C.W., Burlingame A.L.,
RA Barondes S.H., Leflier H.;
RT "Soluble lactose-binding lectin from rat intestine with two different
carbohydrate-binding domains in the same peptide chain.";
RL J. Biol. Chem. 268:5929-5939(1993).

-1- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
SUGARS.
-1- SUBUNIT: Monomer.
-1- TISSUE SPECIFICITY: Highly expressed in full-length form in small
and large intestine and stomach but was not detected in other
tissues including lung, liver, kidney and spleen.
-1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
domains.
-1- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```


CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).

CC -----
DR EMBL: I78132; AAB51605.1; -
DR EMBL: X91790; CAA62904.1; ALT_INIT.
DR EMBL: AF074000; AAD45402.1; -
DR EMBL: AF074001; AAD45403.1; -
DR EMBL: AF074002; AAD45404.1; -
DR EMBL: AF193806; AAF19370.1; ALT_INIT.
DR EMBL: AF193805; AAF19370.1; JOINED.
DR EMBL: AF342815; AAK16735.1; ALT_INIT.
DR EMBL: AL136105; CAC15946.1; ALT_INIT.
DR EMBL: AL136105; CAC15947.1; ALT_INIT.
DR EMBL: BC018618; AAI18618.1; -
DR EMBL: BC016486; AAI16486.1; ALT_INIT.
DR HSSP: P17931; IAK3.
DR Genew; HGNC:6569; LGALS8.
DR MIM: 606099; -
DR GO; GO:0005615; Cextracellular space; TAS.
DR GO; GO:0005530; F:lectin; TAS.
DR InterPro; IPR008985; Cona_like_lect_gl.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind lectin; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
KW Domain; Lectin; Repeat; Antigen; Alternative splicing; Polymorphism.
FT DOMAIN 1 153 GALAPTIN 1.
FT DOMAIN 154 184 LINKER.
FT DOMAIN 185 316 GALAPTIN 2.
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
FT VARSPPLIC 182 182 L -> LPSNRGDISKIAPIPTVTKSKDSVNHNTLCTKIP
FT PMSVSK (in isoform 2).
FT VARSPPLIC 182 182 /FTID=VSP_003094.
FT VARSPPLIC 182 182 L -> LQVSPSPMDLQSGSEFTCSVIMTRVLEIAFCRPI
FT /GALTASRQ (in isoform 3).
FT /FTID=VSP_003095.
FT Y -> F.
FT C -> R.
FT V -> M.
FT /FTID=VAR_009710.
FT V -> M.
FT /FTID=VAR_012991.
FT KRE -> OKER (IN REF. 2).
FT D -> A (IN REF. 2).
FT S -> V (IN REF. 1).
FT S -> R (IN REF. 2, 6, 7 AND 8).
FT K -> Q (IN REF. 1).
FT X -> Q (IN REF. 1).
SQ SEQUENCE 316 AA; 35538 MW; 1724C36085A0510A CRC64;
Query Match 16.8%; Score 228.5; DB 1; Length 316;
Best Local Similarity 40.0%; Pred. No. 9.7e-08;
Matches 56; Conservative 22; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVEYNPLPFGVPRMLITLGTCKPNANRIADFGNG-----DVAHF-PRFENNRR 165
DB 15 VIVYVGTITPDQDGLIIVICGHVPSDADRFQVDLQNGSVKRAVVAHFHNRFRAG- 73
QY 166 RIVYCNKTLNNNGREERQSVFPESGPKPIQVLPDFEFKAVANDAH-LQYNERVKUL 224
DB 74 -CIVCNLTILNEKGMREITITDMPFKREKSPFIYIMVLKQFQVAVANGKILLYGRIGP- 131
QY 225 NEISKIGISGIDILTSASYT 244
DB 132 EKIDTIGIYGVKVINHSIGFS 151
RESULT 15
LEG8 RAT STANDARD; PRT; 316 AA.
AC 062665;
DT 01-NOV-1997 (Rel. 35, Created)
DE 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Galectin-8 (30 kDa S-type lectin) (RL-30).
GN LGALS8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=liver;
RX MEDLINE=95155445; PubMed=7852431;
RA Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y.;
RT Galectin-8, A new rat lectin, related to galectin-4.";
RL J. Biol. Chem. 270:3447-3453(1995).
CC -! FUNCTION: Possesses sugar binding and hemagglutination activity.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! TISSUE SPECIFICITY: Expressed in liver, kidney, cardiac muscle,
CC lung, and brain.
CC -! DEVELOPMENTAL STAGE: Very low levels in whole embryos, high levels
CC in adult tissues.
CC -! DOMAIN: Contains two homologous but distinct carbohydrate-binding
CC domains.
CC -! SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, courtesy of
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).

DR EMBL: U09824; AAB6359.1; -
DR PIR; A55975; A55975.
DR HSSP; P17931; IAK3.
DR InterPro; IPR008985; Cona_like_lect_gl.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
KW Galectin; Lectin; Repeat.
FT DOMAIN 1 153 GALAPTIN 1.
FT DOMAIN 154 184 LINKER.
FT DOMAIN 185 316 GALAPTIN 2.
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 36038 MW; C04B76CFFB913D59 CRC64;
Query Match 16.1%; Score 218.5; DB 1; Length 316;
Best Local Similarity 39.9%; Pred. No. 4e-07;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VEPNPLPFGVPRMLITLGTCKPNANRIADFGNG-----DVAHF-PRFENNRR 166
DB 16 IPEVSTITBQLKFGSLIVIRGHVPSDADRFQVDLQNGSLKPRADVAHFHNRFRKSN-- 73
QY 167 VIVYCNKTLNNNGREERQSVFPESGPKPIQVLPDFEFKAVANDAH-LQYNERVKUL 225
DB 74 CIVCNLTILNEKGMREITITDMPFKREKSPFIYIMVLKQFQVAVANGKILLYAHRLNP- 132
QY 226 EISKIGISGIDILTSASY 243
DB 133 KIDTIGIYGVKVINHSIGF 150

RESULT 16
LEG8 HUMAN STANDARD; PRT; 323 AA.
AC P56470;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE (U36LBP) (Antigen NY-CO-27).

GN LGALS4.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Gastric carcinoma;
 RA Kato S.;
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colorectal carcinoma;
 RA Redreche H., Mallo G.V., Montalco G., Dagorn J.C., Iovanna J.L.;
 RT "Cloning and expression of the mRNA of human galectin-4, an s-type
 RT lectin down-regulated in colorectal cancer.";
 RL Eur. J. Biochem. 248:225-230 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97306339; PubMed=9162064;
 RT Hufelt M.E., Jordan E.T., Gilt M.A., Barondes S.H., Lefler H.;
 RT "Strikingly different localization of galectin-3 and galectin-4 in
 RT human colon adenocarcinoma T84 cells. Galectin-4 is localized at sites
 RT of cell adhesion.";
 RL J. Biol. Chem. 272:14294-14303 (1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon, and Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Ditchenko L., Marusina K., Farmer A.A., Rubin A.A., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshitoki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.N., Gay L.J., Hilyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madañ A.C., Shevchenko Y., Bouffard G.G.,
 RA Whiteing M., Madañ A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -1- FUNCTION: Galectin that binds lactose and a related range of
 CC sugars. May be involved in the assembly of adherens junctions.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
 CC domains.
 CC -1- SIMILARITY: Belongs to the galectin (galaplin/s-lectin) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB006781; BAA22165.1; -;
 DR EMBL; AF014838; AAC51763.1; -;
 DR EMBL; U82953; AAB86590.1; -;
 DR EMBL; BC003661; AAH03661.1; -;
 DR EMBL; BC005146; AAH05146.1; -;
 DR EMBL; BC034750; AAH34750.1; -;
 DR HSSP; P47929; 1BKZ;
 DR Genew; HGNC:6565; LGALS4.

DR MM; 602518; -;
 DR GO; GO:0005829; C:cytosol; TAS.
 DR GO; GO:0005886; C:plasma membrane; TAS.
 DR GO; GO:0005530; F:lectin; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR InterPro; IPR008985; ConA_like_lect_91.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 2.
 DR SMART; SM00276; GLBCT; 2.
 DR PROSITE; PS00309; GALAPTIN; 2.
 KW Galectin; Lectin; Repeat.
 FT DOMAIN 1 152 GALAPTIN 1.
 FT DOMAIN 153 177 LINKER.
 FT DOMAIN 178 323 GALAPTIN 2.
 FT BINDING 256 262 BETA-GALACTOSIDE (BY SIMILARITY).
 SQ SEQUENCE 323 AA; 35941 MW; E79EC0A9AB3990ER CAC64;
 Query Match 15.9%; Score 215.5; DB 1; Length 323;
 Best local Similarity 34.4%; Pred. No. 6.3e-07;
 Matches 52; Conservative 28; Mismatches 62; Indels 9; Gaps 5;
 Cy 99 AYATGPAPAPGLIVPEYNTPLPGVVPFRLITLITGVKPNANITADP---QGNQVYA 154
 Db 2 AYVAPGYQPTVNTPLTPYQPIPGGLNVGMSVYIQVASEHKKPFVNFVWGQDPGSDVA 61
 Cy 155 FHF-PRENNRRRYVTCNTKLDNNWGREERQSVFPFESGKPKFQVLEPPHFKVAVN-D 212
 Db 62 FHFNPFRDGMXK--VVENTIQGKRGSEERSRMPEFKGAPFELVFLAHAYKVVYVGN 119
 Cy 213 AHLQYHRYVKLNFISKLGISGIDITLSAST 243
 Db 120 PFVEYGHRL-PLQWVTHLQVGDGLQSIINF 149
 RESULT 17
 LEG7 HUMAN STANDARD; PRT; 135 AA.
 AC P47929;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Galectin-7 (Gal-7) (HKU-14) (P17) (p53-induced protein 1).
 GN LGALS7 OR PIG1.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 7-19; 75-82; 111-117 AND 120-132.
 RC TISSUE=Epidermis;
 RX MEDLINE=95197604; PubMed=7534301;
 RA Madsen P., Rasmussen H.H., Flint T., Gromov P., Kruse T.A., Honore B.,
 RA Vorum H., Celis J.E.;
 RT "Cloning, expression, and chromosome mapping of human galectin-7.";
 RL J. Biol. Chem. 270:5823-5829 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Epidermis;
 RX MEDLINE=95246905; PubMed=7729568;
 RA Magnaldo T., Bennerd F., Darmon M.;
 RT "Galectin-7, a human 14-kDa S-lectin, specifically expressed in
 RT keratinocytes and sensitive to retinoic acid.";
 RL Dev. Biol. 168:259-271 (1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Ditchenko L., Marusina K., Farmer A.A., Rubin A.A., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hutylk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maria M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RA and mouse cDNA sequences.";
 RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RL [4]
 RP FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21671326; PubMed=11706006;
 RA Kuwabara I., Kuwabara Y., Yang R.Y., Schuler M., Green D.R.,
 RA Zuraw B.L., Hsu D.K., Liu F.T.;
 RT "Galectin-7 (PGI) exhibits pro-apoptotic function through JNK
 RT activation and mitochondrial cytochrome c release.";
 RL J. Biol. Chem. 277:3487-3497 (2002).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=98434364; PubMed=9760227;
 RA Leonidas D.D., Vatzaki E.H., Vorun H., Celis J.E., Madsen P.,
 RA Acharya K.R.;
 RT "Structural basis for the recognition of carbohydrates by human
 RT galectin-7.";
 RL Biochemistry 37:13930-13940 (1998).
 CC -1- FUNCTION: Could be involved in cell-cell and/or cell-matrix
 CC interactions necessary for normal growth control. Pro-apoptotic
 CC protein that functions intracellularly upstream of JNK activation
 CC and cytochrome c release.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear. May be secreted by
 CC a non-classical secretory pathway.
 CC -1- TISSUE SPECIFICITY: Mainly in stratified squamous epithelium.
 CC -1- INDUCTION: By p53.
 CC -1- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; L07769; AAA67899.1; -;
 DR EMBL; U06643; AAA68820.1; -;
 DR EMBL; BC042811; AAH42811.1; -;
 DR PIR; I55469; I55469.
 DR PDB; 1BRZ; 04-NOV-98.
 DR PDB; 2GAL; 04-NOV-98.
 DR PDB; 3GAL; 04-NOV-98.
 DR PDB; 4GAL; 04-NOV-98.
 DR PDB; 5GAL; 04-NOV-98.
 DR Aarhus/Chem-2DPAGE; 17; IEF.
 DR Genew; HGNC:6568; LGAL57.
 DR MIM; 600615; -;
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0005530; F:lectin; TAS.
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 DR InterPro; IPR008985; Coma_like_lect_1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-band_lectin; 1.
 DR SMART; SM00276; GLECF; 1.
 DR PROSITE; PS00109; GALAPTIN; 1.
 KM Galectin; Lactin; Apoptosis; Nuclear protein; 3D-structure.
 FT INIT MET 0 0
 FT BINDING 69 75 BETA-GALACTOSIDE (POTENTIAL).
 FT STRAND 5 8

FT TURN 10 11
 FT TURN 15 16
 FT STRAND 18 25
 FT TURN 27 28
 FT STRAND 31 37
 FT TURN 42 43
 FT STRAND 46 53
 FT TURN 54 57
 FT STRAND 58 65
 FT TURN 66 67
 FT STRAND 68 69
 FT STRAND 73 74
 FT TURN 82 83
 FT STRAND 84 92
 FT STRAND 96 101
 FT TURN 102 103
 FT STRAND 104 110
 FT HELIX 115 117
 FT STRAND 120 125
 FT STRAND 128 135
 SQ SEQUENCE 135 AA; 14944 MD; 4E7CEA54036EF806 CRC64;
 Query Match 15.5%; Score 211; DB 1; Length 135;
 Best Local Similarity 35.3%; Pred. No. 5.3e-07;
 Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;
 QY 114 VRYNPLPGGVPRMILITLITGVKRNARIALDF----QRGNDVAFHF--PRENNRRVY 168
 DB 3 VHKSSLPFGIRPGVLRIRGLVPPNARFVNLLCGEGSDALHFFPRIDTS--V 59
 QY 169 VCNLTLDNNNGEEROSVPEFSGKPKIQVLVEPDHFKVAVNDV-HLYNHRVKLANEI 227
 DB 60 VNSKEGSGWKEERKPGVPFGQPFVLLIASDDGFAVVGDAQYHFRRL-PLAVY 118
 QY 228 SKLGISGIDILTS 240
 DB 119 RLVEVGADYDIDS 131
 RESULT 18
 LE68 MOUSE STANDARD; PRT; 316 AA.
 AC 09Jul15;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DE Galectin-8 (LGALS-8).
 GN LGALS8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RA Maier C., Haussler J., Roesch K., Moschgath E., Haussler J.,
 RA Vogel W.;
 RT "The human LGALS-8 gene: genomic sequence and expression of the
 RT prostate carcinoma tumour antigen (PCTA-1) and the Po6 carbohydrate
 RT binding protein.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Possesses sugar binding and hemagglutination activity.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF218069; AAF27645.1; -;
 DR EMBL: BC040243; AAH40243.1; -;
 DR HSSP: P17931; IAXK.
 DR MGD: MGI:1928481; Tga1s8.
 DR InterPro: IPR008985; Cons_1like_1ec_g1.
 DR InterPro: IPR001079; Galectin.
 DR Pfam: PF00337; Gal-bind_lectin; 2.
 DR SMART: SM00276; GLECT; 2.
 DR PROSITE: PS00309; GALACTIN; 1.
 DR KMW: KMW:153; GALACTIN 1.
 DR FT DOMAIN 1 153 GALACTIN 1.
 DR FT DOMAIN 154 184 LINKER.
 DR FT DOMAIN 185 316 GALACTIN 2.
 DR FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
 DR SEQUENCE 316 AA; 36161 MW; 11A20309AFA52C69 CRC64;
 SQ
 Query Match 15.3%; Score 207.5; DB 1; Length 316;
 Best Local Similarity 38.8%; Pred. No. 1.9e-06;
 Matches 54; Conservative 19; Mismatches 55; Indels 11; Gaps 5;
 QY 113 ITPNNLPDGGVPMPLITITGVKPNMRIAIDFGDN-----DVAHF-PRNNENR 165
 DB 15 IIPYVITTEQLKPSLIVIRGHVDESEFQVDQLNSLKPRADVAFHENRPRKSS- 73
 QY 166 RVIVCNTKLDNNWGEERQSVPEESGPEKIQVYVEDHPRVAVNDAH-LQYNRVKTL 224
 DB 74 -CIVCNLTITQEKMGWEELTYDMPFKKESFELIVMLKXKQVAVNGRHVLLYARISP- 131
 QY 225 NEISKLGISGIDILTSASY 243
 DB 132 EQIDTVGIYKVNHSIGF 150
 RESULT 19
 CM32_HUMAN
 ID CM32_HUMAN STANDARD; PRT; 245 AA.
 AC Q8NDG0; Q96BG5;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein c14orf32.
 GN C14ORF32.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN 1
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX PubMed=12508121;
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
 RA Catolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cnudat C.,
 RA Bruns T., Jallou O., Friedlander L., Samson G., Brothier P.,
 RA Cure S., Seguren B., Aniere F., Samin S., Crespeau H., Abbasi N.,
 RA Alich N., Boscus D., Dickhoff R., Dots M., Dubois I., Friedman C.,
 RA Gouyvenoux M., James R., Madan A., Mailey-Betizda B., Mangent S.,
 RA Martins N., Menard M., Ozlas S., Ratcliffe A., Shaffer T., Trask B.,
 RA Vacherie B., Bellemere C., Belser C., Besnard-Comet M.,
 RA Barcol-Mavet D., Boulard M., Brier-Silla S., Combette S.,
 RA Dufosse-Laurent V., Perron C., Lechaplats C., Louesse C., Muselet D.,
 RA Magdelanet G., Patteu E., Petit R., Siraivain-Trukniewicz P., Tryou A.,
 RA Vega-Carany N., Bataille E., Bluet E., Borelats I., Dubois M.,
 RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,
 RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
 RA Verdier J., Verdier-Discale C., Hiller L.W., Fulton L., McPherson J.,
 RA Marsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,
 RA Ouetier F., Waterston R., Hood L., Weissenbach J.,
 RT "The DNA sequence and analysis of human chromosome 14.";
 RL Nature 421:601-607(2003).
 RX [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hesteh F.,
 RA Datchenko L., Marnins K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SIMILARITY: Belongs to the MISS family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AL834285; CAD38959.1; ALT_INIT.
 DR EMBL: AL119316; -; NOT_ANNOTATED_CDS.
 DR EMBL: BC015621; AAH15621.1; -;
 DR Genew; HGNC:19840; C14orf32.
 FT DOMAIN 27 241 PRO-RICH.
 SQ SEQUENCE 245 AA; 24269 MW; 8B03ADFE581D5C5 CRC64;
 QY Query Match 14.6%; Score 198; DB 1; Length 245;
 Best Local Similarity 39.9%; Pred. No. 5.9e-06;
 Matches 57; Conservative 5; Mismatches 47; Indels 34; Gaps 9;
 QY 1 MADDNFSIHDAL-----SGSGNPNPQGMPCA--WGQNPAGAGGTPGASYPGP 45
 DB 1 MSDEFSLADALPEHSAPKTSVSNTPKQPQ-PQGMPSGNPNWNPNSA-----PSSVPSGLP 54

QY 46 GQAPPAPATP-GQAPPAPATP-GAYPAPATP-GPAPGAPSG---QSPAPCA 99
 DB 55 PSATPSTVFPGAPATMPSPVPTGPPGPPA-----FPSPGSPCPPGPGPYAPTPVPG 110
 QY 100 YATGPYAGAPG---TIVPYNLP 119
 DB 111 GPTGPPTTNPMPPELPPEYVPGAP 133
 RESULT 20
 CM32 MOUSE STANDARD; PRT; 242 AA.
 AC Q8BR93, Q8C90, 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein c14orf32 homolog.
 GN C14ORF32.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cecum, Head, Lung, and Urinary bladder;
 RX MEDLINE=22354683; PubMed=1246851;
 RA Okazaki Y., Futuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Oseko N., Saito R., Suzuki H., Yamana H., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.W.,
 RA Schriml L.M., Kampin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake U.A., Bird D., Brusic V., Chothia C., Corbett L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmer S., Gustinchik S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Mik H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perle G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed U.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sadelain A., Schneider C., Semple C.A., Seton M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=2338257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Stappenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Lonnellund N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultx S.W.,
 RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smailus D.E.,
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -1- SIMILARITY: Belongs to the M18 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AK03615; BAC28392.1; -
 DR EMBL; AK035239; BAC28996.1; -
 DR EMBL; AK048061; BAC32228.1; -
 DR EMBL; AK086896; BAC3759.1; -
 DR EMBL; BC047201; AAH47201.1; -
 DR MGI; MGI:2444022; C130032J12Rik.
 FT DOMAIN 35 238 PRO-RIK.
 SQ SEQUENCE 242 AA; 2385 MW; 3E9989AAAF780BC9 CRC64;
 Query Match 14.6%; Score 197.5; DB 1; Length 242;
 Best Local Similarity 31.7%; Pred. No. 6.2e-06;
 Matches 64; Conservative 9; Mismatches 60; Indels 69; Gaps 11;
 QY 1 MADNPSLHPLSGSNPNP-----QGFGA-WGN---QPAGAGVPGASYG 43
 DB 1 MDEPSLADLPEQSAKPATNTKAGHSQGWSSPWSNPAPRPSGLPSS--A 58
 QY 44 YP-----GQAPPAPGQAPPAPGAYHGAAGV---FGADAPGV-YPGP-----PSGPGA 87
 DB 59 ASTVDFGVPTGMPVPMPTGPPPPPPGSCPPGVPAFAVAPGPGPYA 118
 QY 88 YSSQSPAPGAYATGYPAPAGPLI-----VPYNPLUG 122
 DB 119 TNNMPPELP---RPGAPTPDPAAGSLGFWPMSSGPMWAGIAGHPNMPYRSPGY 173
 QY 123 GVPRMLI---TIGTVKNA 140
 DB 174 PTVPPVSGAPVPWCTVPGGA 195
 RESULT 21
 ANX7 HUMAN STANDARD; PRT; 466 AA.
 AC P20073;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Annexin A7 (Annexin VII) (Synexin).
 GN ANX7 OR ANX7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89264510; PubMed=2542947;
 RA Burns A.L., Magenzko K., Shitvan A., Srivastava M., Rojas E.,
 RA Alijani M.R., Pollard H.B.,
 RT "Calcium channel activity of purified human synexin and structure of
 RT the human synexin gene."
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3798-3802 (1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pituitary;
 RX MEDLINE=2338257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Mang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
RA Stadelson M., Soares M.B., Bonaldo M.F., Caavaant T.L., Schneitz T.E.,
RA Brownstein M., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Calcium/phospholipid-binding protein which promotes
CC membrane fusion and is involved in exocytosis.
CC -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC calcium and phospholipid.
CC -!- SIMILARITY: Contains 4 annexin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL: J04543; AAA36616.1; -
DR EMBL: BC002632; AAH02632.1; -
DR PIR: A54467; L0H07.
DR HSSP: P26256; 1DM5.
DR Gene: HGNC:545; ANXA7.
DR MIM: 186360; -
DR InterPro: IPR001464; Annexin.
DR Pfam: PF00191; annexin; 4.
DR PRINTS: PR00196; ANNEXIN.
DR ProDom: PD000143; Annexin; 4.
DR SMART: SM00335; ANX; 4.
DR PROSITE: PS00223; ANNEXIN; 4.
KW Annexin; Calcium/phospholipid-binding; Repeat.
FT DOMAIN 1 143 REPEAT-RICH REGION.
FT REPEAT 172 232 ANNEXIN 1.
FT REPEAT 244 304 ANNEXIN 2.
FT REPEAT 327 387 ANNEXIN 3.
FT REPEAT 403 463 ANNEXIN 4.
FT DOMAIN 5 20 3 X 5 AA TANDEM REPEATS OF G-Y-P-P-X.
FT REPEAT 5 9 1.
FT REPEAT 10 14 2.
FT REPEAT 16 20 3.
SQ SEQUENCE 466 AA; 50315 MW; 09A6760729D45FCD CRC64;

Query Match 14.5%; Score 196.5; DB 1; Length 466;
Best Local Similarity 38.8%; Pred. No. 1.3e-05;
Matches 57; Conservative 6; Mismatches 45; Indels 39; Gaps 11;
QY 17 ENPQGWPGAMGNQAPAGAGYPGASYPGYPGQAPPGAYPGQAPPGAYHGAAPYFGAPGAPG 76
Db 33 PYPGSGP-----PMGAGAYPQVPSGYPG--AGGYA---PGGY-PAPGGYFGAPPG 79
QY 77 ---YYPGPGG-----PGAYPSSQGPAPG-AIATGPy-----GADAGPLIVPNLPL 120
Db 80 GAPSPYGVPGQGVPGGAGVPGGAGFGSGYPPSPSQSYXGSGPAQVPLPGGFGGQMPGQY----- 135
QY 121 PGG--VVPRLITLI---GTVKPYAN 141
Db 136 PGGQPTIPSPATVIVTQVGTIRPAN 162

RESULT 22
ID CAL3_BOVIN STANDARD; PRT; 1049 AA.
AC P04258;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain.
GN COL3A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-242.
RX MEDLINE=80026026; PubMed=488906;
RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wachter E.,
RT Kuhn K.;
RT "The covalent structure of calf skin type III collagen. I. The amino
RT acid sequence of the amino terminal region of the alpha 1(III) chain
RT (positions 1-222).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820 (1979).
RN [2]
RP SEQUENCE OF 243-422.
RX MEDLINE=80026027; PubMed=488907;
RA Dewes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. II. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT (positions 223-402).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832 (1979).
RN [3]
RP SEQUENCE OF 423-571.
RX MEDLINE=80026028; PubMed=488908;
RA Bentz H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. III. The
RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT (positions 403-551).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840 (1979).
RN [4]
RP SEQUENCE OF 572-808.
RX MEDLINE=80026029; PubMed=488909;
RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. IV. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT (positions 552-788).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850 (1979).
RN [5]
RP SEQUENCE OF 809-947.
RX MEDLINE=80026030; PubMed=488910;
RA Dewes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. V. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT (position 789-927).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860 (1979).
RN [6]
RP SEQUENCE OF 948-1049.
RX MEDLINE=80026031; PubMed=488911;
RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. VI. The amino
RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT 1(III)CB9B (positions 928-1028).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868 (1979).
RN [7]
RP FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylysines.
CC -!- PIM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
DR PIR: A02862; CGB07S.

```

DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; C1g helix; 3.
DR PROSITE; PS01208; WFGC_1; PARTIAL.
KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen.
FT DOMAIN 1 14
FT DOMAIN 15 1040
FT DOMAIN 1041 1049
FT MOD_RES 95 95
FT MOD_RES 107 107
FT MOD_RES 119 119
FT MOD_RES 938 938
FT MOD_RES 950 950
FT CARBOHYD 107 107
FT CARBOHYD 950 950
FT DISULFID 1040 1040
FT DISULFID 1041 1041
SQ SEQUENCE 1049 AA; 93651 MW; 8EEC3D1C66EC9A3 CRC64;

Query Match 14.2%; Score 193; DB 1; Length 1049;
Best Local Similarity 33.7%; Pred. No. 4.5e-05;
Matches 64; Conservative 8; Mismatches 58; Indels 60; Gaps 12;

QY 5 FSLHDLG-----SGNNPQGWPGAMGNQAGAGYRGAYPGYRG--QAPRGAYPGQA 57
DB 2 YRAYDVKSSVAGSGIANGYGPAGPPGP--PGTSGHGA--PGAPGYGPPGE--PGQA 56
QY 58 PGAYHAGYAGAY-----PGAPAGYVYGP--SGPGAYP----- 89
DB 57 GPAGPPPGPAGCGPSKQDSGRPGRGFRPPGCMGPGAMPFPFGMKHRRGDGN 116
QY 90 -SSGQPSAPGAY---ATGPGYAPA--GPLIVPNYLPFGVVPRLITLIGTVKPMNR 142
DB 117 GKRGEGAGLKGNGVPGEDGAPGMBGRGAPGRRGPG-----LPQAAGARGN- 166
QY 143 IALDPQRGND 152
DB 167 ---DGARGSD 173

RESULT 23
ANX7_MOUSE STANDARD; PRT; 463 AA.
ID ANX7_MOUSE
AC Q07076;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Annexin A7 (Annexin VII) (Synexin).
GN ANXA7 OR ANX7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93168121; PubMed=7916616;
RA Zhang-Keck Z.Y., Burns A.L., Pollard H.B.;
RT "Mouse synexin (annexin VII) polymorphisms and a phylogenetic
RT comparison with other synexins."
RL Biochem. J. 289:735-741(1993).
CC -1- FUNCTION: Calcium/phospholipid-binding protein which promotes
CC membrane fusion and is involved in exocytosis.
CC -1- DOMAIN: A pair of annexin repeats may form one binding site for
CC calcium and phospholipid.
CC -1- SIMILARITY: Belongs to the annexin family.
CC -1- SIMILARITY: Contains 4 annexin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; I13129; AAA37238.1; -.
DR HSSP; P26256; 1DM5.
DR MGD; MGI:88031; Anxa7.
DR InterPro; IPR001464; Annexin.
DR Pfam; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
DR ProDom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 4.
KM Annexin; Calcium/phospholipid-binding; Repeat.
FT DOMAIN 1 143
FT REPEAT 169 229
FT REPEAT 241 301
FT REPEAT 324 384
FT REPEAT 400 460
FT DOMAIN 5 20
FT REPEAT 5 9
FT REPEAT 10 14
FT REPEAT 16 20
SQ SEQUENCE 463 AA; 49939 MW; 50F7B20FD48EB0D5 CRC64;

Query Match 14.2%; Score 192.5; DB 1; Length 463;
Best Local Similarity 33.0%; Pred. No. 2.3e-05;
Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

QY 11 ISGSGNPPQGWPGAMGNQAPAG-----AGYPGASVYGYGQAP--GAYPGQAPG 60
DB 1 MSYPPYP-PTGYPPPPGYPAGQSSFTAGYD-----YSGPPEMGAYP--PAPSG 52
QY 61 AYHAGPAGY---GAPAGYVYRG--PPSGPAGYPS-----SGQPSAG-AYA 101
DB 53 GYPGA-GGYPAPGAYPAGYRGALSPGPPAYPGQGPGAGGFGYQPPAGSYG 111
QY 102 TGEY-----GAPAGPLIVPNYLPFGGVV-----RMLTILGTVKPMNRIMALDFOR 149
DB 112 GGPAGVVPVPGFPFGGQMSQY---PGGAPYPPSQPASMTOGTGTLIPASN---FDAMR 164
QY 150 GNDVAFHPPRFENNRRVYVCTKLDNNWGREROSVFPFESGK 193
DB 165 DAELIRKAMKGFCTDEQALVDVVSNRSDQRIIAAFITWTK 208

RESULT 24
LE32_CAEEL STANDARD; PRT; 279 AA.
ID LE32_CAEEL
AC P36573;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 32 kDa beta-galactoside-binding lectin (32 kDa GBP).
GN W09H1.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348399; PubMed=1639789;
RA Hirabayashi J., Satoh M., Kasai K.-I.;
RT "Evidence that Caenorhabditis elegans 32-kDa beta-galactoside-binding
RT protein is homologous to vertebrate beta-galactoside-binding lectins.
RT cDNA cloning and deduced amino acid sequence."
RL J. Biol. Chem. 267:15485-15490(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476274; PubMed=9334250;
RA Arata Y., Hirabayashi J., Kasai K.-I.;
RT "Structure of the 32-kDa galectin gene of the nematode Caenorhabditis

```

RL elegans.";
 RL J. Biol. Chem. 272:26669-26677(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Brictol N2;
 RL Smye R.;
 RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 202-220.
 RX MEDLINE=92348337; PubMed=1639749;
 RA Hirabayashi J., Satoh M., Ohnaya Y., Kasai K.-I.;
 RT "Purification and characterization of beta-galactoside-binding
 RT proteins from Caenorhabditis elegans.";
 RL J. Biochem. 111:553-555(1992).
 CC -1- FUNCTION: Binds galactose.
 CC -1- SIMILARITY: Belongs to the galectin (galactin/s-lectin) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL; M94671; AAB87718.1; -;
 DR EMBL; AB000802; BAA22942.1; -;
 DR EMBL; Z82081; CAB04959.1; -;
 DR PIR; PX0062; PX0062;
 DR PIR; T37216; T37216;
 DR HSSP; P17931; 1A3K.
 DR WormPep; W09H1.6a; CE16576.
 DR InterPro; IPR008985; ConA like_1ec_g1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 2.
 DR SMART; SM00276; GLECT; 2.
 DR PROSITE; PS00309; GALAPTIN; 2.
 DR Galectin; Lectin; Repeat; Multigene family.
 DR DOMAIN 1 146 GALAPTIN 1.
 DR DOMAIN 147 279 GALAPTIN 2.
 FT BINDING 213 219 BETA-GALACTOSIDE (BY SIMILARITY).
 SO SEQUENCE 279 AA; 31809 MW; ED9AE2A837571DA CRC64;
 Query Match 13.9%; Score 188.5; DB 1; Length 279;
 Best Local Similarity 36.4%; Pred. No. 2.6e-05;
 Matches 47; Conservative 26; Mismatches 49; Indels 7; Gaps 5;
 QY 114 VPKLPLPGVVPKMLITITLTKVKNRRLALDFQRGN-DVAHPF-PFENNRRIVICN 171
 DB 150 VPKSGLANGLPVGSKLIVFGEVKKARFHFVNLIRKNGDISFHNPFDEKH--VLRN 206
 QY 172 TKLNNMGREROSVFPESGPKFQVLVEPDHPKXAVN-DALHQNHRKXKLNESKL 230
 DB 207 SLANWENREKGNPKPKVGVPLVQNEEYAVQVFNNGERYISFPHRDP-HDIAGL 265
 QY 231 GISGDIDLT 239
 DB 266 QISGDIELS 274
 RESULT 25
 CBA DICTI STANDARD; PRT; 467 AA.
 AC P35085;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Calcium-binding protein.
 GN CBP OR CBPA.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 CC NCBI_TaxID=44689;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX2;
 RL Wenington R., Greenwood M., Tsang A.;
 RA Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL; U03413; AAA03471.1; -;
 DR DictyBase; DDB0002101; cbp.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR006031; XYPPX.
 DR Pfam; PF00036; efhand; 2.
 DR Pfam; PF02162; XYPPX; 18.
 DR SMART; SM00054; EFH; 2.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR Calcium-binding.
 KW CA BIND 412 423 EF_HAND (POTENTIAL).
 SO SEQUENCE 467 AA; 48976 MW; 2D2055D1E344FC18 CRC64;
 Query Match 13.9%; Score 188; DB 1; Length 467;
 Best Local Similarity 37.2%; Pred. No. 4.4e-05;
 Matches 54; Conservative 0; Mismatches 35; Indels 56; Gaps 8;
 QY 12 SGSGNPNGQWPGAMGNOPAGAGGYP-----GASYPGPGAP-----GAYPGAP 58
 DB 113 SGQYPPQPGQPGYPPQPGAPGYPQPGQPGQPGQPGQPGQPGQPGQPGQPG 172
 QY 59 -PGAY-----HGAPAY-----PGAPAP-GVYPGPPSGP----- 85
 DB 173 QPGAYPPQPGQPGQPGAYPPQPGQVQNTLAKTGAQPGVPPPGAYPGQPGVPPQPG 232
 QY 86 ----GAYPSSGQPSA-----PGAY 100
 DB 233 QPPWAGYPPQPGQPGAYPPQPGQPGAY 257
 RESULT 26
 LEG1 HAECO STANDARD; PRT; 283 AA.
 ID LEG1 HAECO
 AC 044126;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 32 kDa beta-galactoside-binding lectin (galectin 1).
 GN GAL-1.
 OS Haemonchus contortus (Barber pole worm).
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 CC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
 CC NCBI_TaxID=6289;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Moredu; TISSUE=Gut;
 RX MEDLINE=20067137; PubMed=10599080;
 RA Newlands G.F.J., Skuce P.J., Knox D.P., Smith S.K., Smith W.D.;
 RT "Cloning and characterization of a beta-galactoside-binding protein
 RT (galectin) from the gut of the gastrointestinal nematode parasite
 RT Haemonchus contortus.";
 RL Parasitology 119:483-490(1999).
 CC -1- FUNCTION: Binds galactose.
 CC -1- SIMILARITY: Belongs to the galectin (galactin/s-lectin) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L41162; AAC1947.1; -

DR EMBL; AF026802; AAD47357.1; -

DR EMBL; AF026801; AAD47357.1; JOINED.

DR EMBL; AL035669; CAC12750.1; -

DR EMBL; BC011705; AAH11705.1; -

DR EMBL; X91013; CAA62495.1; -

DR Genew; HGNC:2219; COL9A3.

DR MIM; 120270; -

DR MIM; 600969; -

DR GO; 0005594; C:collagen type IX; TAS.

DR InterPro; IPR008161; Clg_helix.

DR InterPro; IPR008160; Collagen.

DR Pfam; PF01391; Collagen; 9.

DR ProDom; PD000007; Clg_helix; 3.

KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Signal; Polymorphism.

FT SIGNAL 1 25

FT CHAIN 26 684

FT DOMAIN 29 519

FT DOMAIN 520 550

FT DOMAIN 551 630

FT DOMAIN 631 632

FT DOMAIN 633 661

FT DOMAIN 662 684

FT SITE 423 425

FT SITE 603

FT CARBOYD 483

FT VARIANT 563

FT VARIANT 565

FT VARIANT 564

FT VARIANT 566

FT CONFLICT 18 18

FT CONFLICT 39 39

FT CONFLICT 144 144

FT CONFLICT 435 435

FT CONFLICT 524 524

FT CONFLICT 576 576

SEQ SEQUENCE 684 AA; 63616 MW; 892F035CE6B6733 CRC64;

Query Match 13.8%; Score 187; DB 1; Length 684;

Best Local Similarity 32.0%; Pred. No. 7.2e-05;

Matches 57; Conservative 13; Mismatches 60; Indels 48; Gaps 8;

QY 9 DALSG-SGNPNPGQWPGAMGN-QPAGAGYFGAYPGYPGA---PPGAYPGQAPPGAY 62

DB 85 DGLTGRDGPPEPKAPGERSLSLGGPPGLGGKGLPGPPEBAGVSGPPGIGLRGPPG-- 142

QY 63 HGAPGAYGAPAPGVYGPSPGAPYSS-----GQSPAPGAYATGPGY- 106

DB 143 ---PSGLPGLRGPFGPPGPHGVLEGGATDLCPSICPPGPPGPMG--FKGPTGY 197

QY 107 -----APAGLIVPNLPLPGVVPMLITIGTVKPNMRLALDQ 148

DB 198 KREGGEVGDKEKDPGPBPAGLPGSVGLG---PRLGLPLPGLPPGDRGPIGR 252

RESULT 28

CA18_HUMAN STANDARD; PRT; 744 AA.

AC P27658; Q96D07;

DT 01-AUG-1992 (Rel. 23, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Collagen alpha 1 (VIII) chain precursor (Endothelial collagen).

GN COL8A1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91231001; PubMed=2029894;

RA Muragaki Y., Mattei M.-G., Yanaguchi N., Olsen B.R., Nimomiya Y.,

RT "The complete primary structure of the human alpha 1 (VIII) chain and

RL assignment of its gene (COL8A1) to chromosome 3.";

RN Eur. J. Biochem. 197;615-622(1991).

[2]

RP SEQUENCE FROM N.A.

PC TISSUE=lung;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Ditschenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,

RA Villalon D.K., Muzny K.D., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,

RA Schnerch A., Schein J.E., Jones S.U.M., Matra W.A.,

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99;16899-16903(2002).

CC -1- FUNCTION: Major component of the Descemet's membrane (basement

CC membrane) of corneal endothelial cells.

CC -1- SUBUNIT: May form homotrimers, or heterotrimers in association

CC with alpha 2 (VIII) type collagens.

CC -1- PTM: Prolines at the third position of the tripeptide repeating

CC unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -1- MISCELLANEOUS: Four consecutive Gly-Pro-Pro triplets are present

CC at the C-terminus of the triple-helical region. These may provide

CC the high thermal stability of this region.

CC -1- SIMILARITY: STRONG, 10 ALPHA 2 TYPES VIII AND X COLLAGENS.

CC -1- SIMILARITY: Contains 1 Clq domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X57527; CAA40748.1; -

DR EMBL; BC013581; AAH13581.1; -

DR PIR; S15435; S15435.

DR Genew; HGNC:2215; COL8A1.

DR MIM; 120251; -

DR GO; 0005591; C:collagen type VIII; TAS.

DR InterPro; IPR001073; Clq.

DR InterPro; IPR008161; Clq_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR008983; TNF_like.

DR Pfam; PF00386; Clq; 1.

DR Pfam; PF01391; Collagen; 8.

DR PRINTS; PR00007; COMPLEMENTC1Q.

DR ProDom; PD000007; Clq_helix; 2.

DR SMART; SM00110; Clq; 1.

DR PROSITE; PS01113; Clq; 1.

DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KM Glycoprotein; Cell adhesion; Collagen; Signal.

FT SIGNAL 1 28

FT CHAIN 29 744

FT DOMAIN 29 117

FT DOMAIN 118 571

COLLAGEN ALPHA 1 (VIII) CHAIN.

NONHELICAL REGION (NC2).

TRIPLE-HELICAL REGION (COL1).

```

FT DOMAIN 572 744 NONHEMICAL REGION (NCL).
FT 609 744 C1Q.
FT CONFLICT 262 262 P -> L (IN REF. 1).
FT CONFLICT 297 297 P -> R (IN REF. 1).
FT CONFLICT 344 344 P -> A (IN REF. 1).
FT CONFLICT 382 382 P -> S (IN REF. 1).
FT CONFLICT 388 388 P -> S (IN REF. 1).
FT CONFLICT 454 454 L -> F (IN REF. 1).
FT CONFLICT 464 464 A -> H (IN REF. 1).
FT CONFLICT 601 601 Y -> T (IN REF. 1).
FT CONFLICT 631 631 A -> G (IN REF. 1).
SQ SEQUENCE 744 AA; 73364 MW; 2BC1B0955DE2C9A3 CRC64;

Query Match 13.6%; Score 185; DB 1; Length 744;
Best Local Similarity 35.0%; Pred. No. 0.0001;
Matches 55; Conservative 8; Mismatches 60; Indels 34; Gaps 7;

QY 15 GNPNGQPMGAMGNC-PAGAGGYPG-----ASYGYGPG-----QAPGAYPGQAP 59
DB 447 GPPGKRLGPIPGKGEAGQKVPGLPGVPLGLGPKGEPGIPGDOGLOGPFG-IPGIGP 505
QY 60 GAYHGAPE-----AYGAPAPGVYP-----GPPSGGAYSSGQSPAPGAYATG 103
DB 506 SGPIPGPIPGPKGEPGLPGPFGPGIGKPGVAGLHGPKPGKALGPQGQPLFG--PPG 563
QY 104 PYGAPAGLIVPYNLPLPGGVPERMLITILGTVKNA 140
DB 564 PPGPPGPAVMPPTPPQGEYLIDMGLGIDGVKPPHA 600

RESULT 29
ID LEG7_RAT STANDARD; PRT; 135 AA.
AC P97590; 054958;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-7 (Gal-7).
GN LGALS7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RA In J.X.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.
MEDLINE=98362649; PubMed=9697310;
RA Magnaldo T., Fowles D., Darmon M.;
RT "Galectin-7, a marker of all types of stratified epithelia.";
RL Differentiation 63:159-168(1998).
CC - FUNCTION: Could be involved in cell-cell and/or cell-matrix
CC interactions necessary for normal growth control. Pro-apoptotic
CC protein that functions intracellularly upstream of JNK activation
CC and cytochrome c release (By similarity).
CC - SUBUNIT: Monomer (By similarity).
CC - SUBCELLULAR LOCATION: Cytoplasmic and nuclear. May be secreted by
CC a non-classical secretory pathway (By similarity).
CC - SIMILARITY: Belongs to the galectin (galactin/S-lec) family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67883; AAB40658.1; -

```

```

DR EMBL; AF036941; AAB88871.1; -.
DR HSSP; P47929; 2GAL.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_Lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
DR Galectin; Lectin; Apoptosis; Nuclear protein.
FT INIT MET 0
FT BINDING 69 75
FT CONFLICT 46 46 BETA-GALACTOSIDE (POTENTIAL).
FT CONFLICT 1 4 SATL -> M (IN REF. 2).
SQ SEQUENCE 135 AA; 15154 MW; FE6215D31944CBAD CRC64;

Query Match 13.6%; Score 184; DB 1; Length 135;
Best Local Similarity 32.8%; Pred. No. 2.5e-05;
Matches 43; Conservative 26; Mismatches 52; Indels 10; Gaps 5;

QY 116 YNLPLPGGVPRMLITILGTVKPNANRIALDPKGN-----DVAFPF-PFENNRRVIVC 170
DB 5 HKTLPQGVRLGTVMRIKGVPPDQAGRHWLNLGCEBGEADALHFNRLDTSS---VVF 61
QY 171 NTKLDNMWGREERQSVPPESGKPKIQLVLPDHFQVAV-NDAAHQNNHRYKXKLNELSK 229
DB 62 NTKQQKWKREERGTGLPFQRGQFEVLITTEEGFTVIGDDEYIHPHHRPSSN-VRS 120
QY 230 LGISGIDLTLS 240
DB 121 VEVGGDVQLHS 131

RESULT 30
ID CAl8_MOUSE STANDARD; PRT; 743 AA.
AC Q00780; Q9D2V4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(VIII) chain precursor.
GN COL8A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAUB/c;
RA Muragaki Y., Shiota C., Inoue M., Ooshima A., Olsen B.R.,
RA Niinomiya Y.;
RT "Alpha 1(VIII)-collagen gene transcripts encode a short-chain
RT collagen polypeptide and are expressed by various epithelial,"
RT endothelial and mesenchymal cells in newborn mouse tissues.";
RL Eur. J. Biochem. 207:895-902(1992).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=C57BL/6J; TISSUE=Kidney;
CC MEDLINE=21085660; PubMed=11217851;
CC Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
CC Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi U., Fukuda S.,
CC Arakawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
CC Saito T., Okazaki Y., Gojobori T., Bono H., Kaetzkawa T., Saito R.,
CC Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
CC Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
CC Kuehl P., Lewis S., Matsuo Y., Nikiado I., Pesole G., Quackenbush J.,
CC Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
CC Sakai K., Okido T., Furuno M., Aono H., Baladrelli R., Barish G.,
CC Blake U., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
CC Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
CC Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
CC Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombaerts P.,
CC Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
CC Sasaki H., Sato K., Schoenbach C., Seya T., Shihata Y., Storch K.-F.,

```



```

CC      in adipose tissue.
CC      -1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
CC      domains.
CC      -1- SIMILARITY: Belongs to the galectin (galactin/s-lectin) family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF244974; AAK77328.1; -
DR      EMBL; AF244975; AAK77329.1; -
DR      EMBL; AF244976; AAK77330.1; -
DR      EMBL; AF244977; AAK77331.1; -
DR      EMBL; AF310686; AAG40863.1; -
DR      EMBL; AF310687; AAG40864.1; -
DR      EMBL; AF222695; AAF34677.1; -
DR      EMBL; BC028222; AAH28222.1; -
DR      Genew; HGNC:15786; LGALSI2.
DR      MIM; 606096; -
DR      InterPro; IPR008985; Cona_like_lect_gl.
DR      InterPro; IPR01079; Galectin.
DR      Pfam; PF00337; Gal-bind_lectin; 1.
DR      SMART; SM00276; GLECT; 1.
DR      PROSITE; PS00309; GALAPTIN; FALSE_NEG.
DR      KMW; Galectin; Lectin; Repeat; Nuclear protein; Alternative splicing.
FT      DOMAIN 47 183
FT      VARSPLIC 194 336
FT      VARSPLIC 1 22
FT      VARSPLIC 1 61
FT      VARSPLIC 74 74
FT      VARSPLIC 200 208
FT      VARSPLIC 336 AA; 37541 MW; F35D0AOCES09E795 CRC64;
SQ      SEQUENCE
Query Match 13.3%; Score 181; DB 1; Length 336;
Best Local Similarity 31.0%; Pred. No. 8.8e-05;
Matches 58; Conservative 26; Mismatches 85; Indels 18; Gaps 7;
QY 73 PAPGYPPGSPGAGVPPSSGQPS-----APGAYATPYPAPAGPLVPPNLPDPGQVVR 127
DB 4 PGGGRAPGTRIVSWSCPTVWSPGKLDPIPSDFIIQP---PVFHPVPVPTTIFGGLHAG 60
QY 128 MTTITIGTVKPNANRIALDFORG-----NDVAFHF--PREFENNRREIVYCNKTLDNMGR 180
DB 61 KVMVLGVVPLDHRRCVDFQCCSCRPRLIAFHNPFFHTTKHVI--CNLHGRMR 119
QY 181 EERQSVFPESGKPKFIQVLVEPDHFKVANDAH--LQVNHRYKLNLSKLGISGIDIT 239
DB 120 EARMWPLALRRSSFLITFLFGNEEVKVSNGCHLHFYRL--PLSHVDTLGIIFDILVE 178
QY 240 SASYTM 246
DB 179 AVGFINT 185
RESULT 32
CA18_RABIT STANDARD; PRT; 744 AA.
AC P14282;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).
GN COL8A1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

CC      Mammalia; Eutheria; lagomorpha; Leporidae; Oryctolagus.
CC      NCBI_TaxID=9986;
CC      RN
CC      RN (1)
CC      SEQUENCE FROM N.A.
RX MEDLINE=89380199; PubMed=2476437;
RA Yamaguchi N., Banya P.D., van der Rest M., Ninomiya Y.;
RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs
RT demonstrate that type VIII collagen is a short chain collagen and
RT contains triple-helical and carboxyl-terminal non-triple-helical
RT domains similar to those of type X collagen.";
RL J. Biol. Chem. 264:16022-16029(1989).
CC      -1- FUNCTION: Major component of the Descemet's membrane (basement
CC      membrane) of corneal endothelial cells.
CC      -1- SUBUNIT: May form homotrimers, or heterotrimers in association
CC      with alpha 2(VIII) type collagens.
CC      -1- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC      -1- MISCELLANEOUS: 4 consecutive G-P-P tripeptides are present at the
CC      C-terminus of the triple-helical region. These may provide high
CC      thermal stability of this region.
CC      -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC      -1- SIMILARITY: Contains 1 C1Q domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J05042; AAA31204.1; -
DR      PIR; A34246; A34246.
DR      InterPro; IPR001073; C1Q.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR008983; TNF_like.
DR      Pfam; PF00386; C1Q; 1.
DR      Pfam; PF01391; Collagen; 8.
DR      PRINTS; PR00067; COMPLEMENTC1Q.
DR      SMART; SM00110; C1Q; 1.
DR      PROSITE; PS01113; C1Q; 1.
DR      KMW; Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
DR      Glycoprotein; Cell adhesion; Collagen; Signal.
FT      SIGNAL 1 20
FT      CHAIN 21 744
FT      DOMAIN 29 117
FT      DOMAIN 118 571
FT      DOMAIN 572 744
FT      DOMAIN 609 744
FT      DOMAIN 744 AA; 73358 MW; 2A8CEFLF8274E99 CRC64;
SQ      SEQUENCE
Query Match 13.3%; Score 180; DB 1; Length 744;
Best Local Similarity 32.4%; Pred. No. 0.00021;
Matches 59; Conservative 10; Mismatches 67; Indels 46; Gaps 10;
QY 15 GNPNDQWPGWAGNPDAGAGYPG--ASYYPGYG-----CAPPGAYPGQA 57
DB 447 GPPGIRGLDPRIC--PKGAGHGKGLGLPGVYGLLGPKEGPIPDQGLQGPPG--IPGIT 503
QY 58 PPGAYHAGP-----AYGAPAPGVY-----GPPSGPGAYPSSGQPSAPGAYA 101
DB 504 GPSGPIGPPGIPGPKGEPLPGPPGPGVGVKPGVAGLHGPSPGALGPQGPGLPG--P 561
QY 102 TGPYAPAGPLLVPNLP--GGVVPMLTTITIGTVKPNANRIALDFORGANDVAFHPP 158
DB 562 PGPBPBPBPVAVP--PTPAPQGYLPDMGLIDGVKTIPIA--YAAKKKGKGGPAYEMP 616
QY 159 RF 160
DB 617 AF 618
RESULT 33

```

```

CC12_CABEL
ID _CC12_CABEL STANDARD; PRT; 316 AA.
AC P20630;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cuticle collagen 12 precursor.
GN COL-12 OR F15H10.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90172409; PubMed=1689778;
RA Park Y.-S., Kramer J.M.;
RT "Randomly duplicated Caenorhabditis elegans collagen genes differ in
RT their modes of splicing."
RL J. Mol. Biol. 211:395-406(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC CC
CC -1- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -1- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -1- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X51622; CA35954.1; -.
DR EMBL; Z73972; CA98257.1; -.
DR PIR; S08169; S08169.
DR WormPep; F15H10.1; CE05638.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 3.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL.
FT CHAIN 1 36 POTENTIAL.
FT DOMAIN 37 316 CUTICLE COLLAGEN 12.
FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
FT DOMAIN 240 266 TRIPLE-HELICAL REGION.
FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
SQ SEQUENCE 316 AA; 30098 MW; 6CA81FP94706D42E CRC64;

Query Match 13.2%; Score 179; DB 1; Length 316;
Best Local Similarity 35.5%; Pred. No. 0.00011;
Matches 55; Conservative 4; Mismatches 50; Indels 46; Gaps 8;

13 GSGNPNQGMWPGMNGNPAGAGYPGASYPGPGQ-----APPG-- 51
Db 126 GSGAGAGAGSPGQDG--APGNDGAPGA--FGNPGQDASEDTAGPDSFCPCPAGPSPGS 181
QY 52 AAYGQAPPGAYHGAPG-----AYGAPAPGVYPPGPGAPYSSGQSPAPGAY-----A 101
Db 182 GAGGQKPGSPGAPGAGPAGGAGALFPGPPGAPGPGAPGAGPAGPAGPAGVVDVPGT 241
QY 102 TGPYGAAPGFLIVPYNLP-----LPGGVYPR 127

```

```

Db 242 PGAPGPGSP--GPAGAPGPGAGGAGSSGPGGPGPG 274
RESULT 34
CC13_CABEL
ID _CC13_CABEL STANDARD; PRT; 316 AA.
AC P20631;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cuticle collagen 13 precursor.
GN COL-13 OR F15H10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90172409; PubMed=1689778;
RA Park Y.-S., Kramer J.M.;
RT "Randomly duplicated Caenorhabditis elegans collagen genes differ in
RT their modes of splicing."
RL J. Mol. Biol. 211:395-406(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC CC
CC -1- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -1- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -1- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X51623; CA35955.1; -.
DR EMBL; Z73972; CA98258.1; -.
DR PIR; S08170; S08170.
DR WormPep; F15H10.2; CE05639.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 3.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL.
FT CHAIN 1 36 POTENTIAL.
FT DOMAIN 37 316 CUTICLE COLLAGEN 13.
FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
FT DOMAIN 240 266 TRIPLE-HELICAL REGION.
FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
SQ SEQUENCE 316 AA; 30100 MW; 00C6D08BEC4701AF CRC64;

Query Match 13.2%; Score 179; DB 1; Length 316;
Best Local Similarity 35.5%; Pred. No. 0.00011;
Matches 55; Conservative 4; Mismatches 50; Indels 46; Gaps 8;

13 GSGNPNQGMWPGMNGNPAGAGYPGASYPGPGQ-----APPG-- 51
Db 126 GSGAGAGAGSPGQDG--APGNDGAPGA--FGNPGQDASEDTAGPDSFCPCPAGPSPGS 181
QY 52 AAYGQAPPGAYHGAPG-----AYGAPAPGVYPPGPGAPYSSGQSPAPGAY-----A 101
Db 182 GAGGQKPGSPGAPGAGPAGGAGALFPGPPGAPGPGAPGAGPAGPAGPAGVVDVPGT 241
QY 102 TGPYGAAPGFLIVPYNLP-----LPGGVYPR 127

```



```

DR Pfam: PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
DR Prodom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 3.
KM Annexin; Calcium/phospholipid-binding; Repeat.
FT REPEAT 218 277 ANNEXIN 1.
FT REPEAT 290 350 ANNEXIN 2.
FT REPEAT 373 433 ANNEXIN 3.
FT REPEAT 449 509 ANNEXIN 4.
SQ SEQUENCE 512 AA; 53314 MW; 70D532D2524388A4 CRC64;

Query Match 12.8%; Score 173.5; DB 1; Length 512;
Best Local Similarity 31.6%; Pred. No. 0.00038;
Matches 73; Conservative 16; Mismatches 71; Indels 71; Gaps 17;

QY 12 SGGNGNPNQ--GMPGAMGNOPAGAGYPGA--SYPGYSG-----QAPPGA-----Y 53
DB 48 SGGGNAPPGGGYPPHA--GGYPA--PGYTGCMPSYFGAGFGAPAGAGGYPGAPAYGV 105
QY 54 PGQAPGAYHGAHPAYPGAPAPGVTPPPSG---PGAYPSG---QPS----- 95
DB 106 PGYGGPFG--FNAPAGGY--GAPNAGCGVPAGGSGPGAGPGYGGFSGSSQSGYAGGPGQ 163
QY 96 -----APGAYATGAYGAPAGAPLIVPYNLPDGGVVRMLITLIGVKNARI 143
DB 164 MPGQMPGMPGAPSGYPSGP--APPAQ--TPY-----AAAMTATGGITKAPNFD 210
QY 144 ALDPQRNDVAFHPPFPENNRRV--IVCNTKLDNNMGREERQSVPPFESGK 193
DB 211 AL--SDAEKLRKAMKMGFTGDEKIDIVVANRSDQ--RQKIGAPKTYAGK 256

RESULT 39
CAS4 CANFA STANDARD; PRT; 754 AA.
AC Q28247;
ID 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; tISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng X., Thorner P.S., Marrano P., Bauman R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV."
RT Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
RL -I- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -I- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -I- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -I- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -I- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -I- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of

```

```

CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -I- DISEASE: A defect in COL4A5 has been found to be the cause of
CC canine X-linked hereditary nephritis (HN), a disease similar to
CC that in humans (also referred to as Alport syndrome) characterized
CC by progressive renal failure and neurosensory deafness.
CC -I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-bib.ch).
CC -----
DR EMBL; U07888; AAB60258.1; -.
DR PIR; A55267; A55267.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 8.
DR Prodom; PD000007; Clg helix; 1.
DR Prodom; PD003923; Procollagen4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON TER 1 1
FT DOMAIN 531 530 TRIPLE-HELICAL REGION.
FT 531 >754 NONHELICAL REGION (NC1).
FT DISULFID 552 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 ? OR 754 (BY SIMILARITY).
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON TER 754 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 12.8%; Score 173.5; DB 1; Length 754;
Best Local Similarity 40.6%; Pred. No. 0.00054;
Matches 52; Conservative 2; Mismatches 61; Indels 13; Gaps 5;

QY 11 LSGSGNPNQMPGAMGNGO-----PAGAGYPGAASYGYGQAPRG--APPGAPPGA 61
DB 227 VGGGGRPWPVPGPPGSKGNPGQDGLPDPAGQKGPQPGFGLPG--PPGLPLSGGKGDG 284
QY 62 YHGAPGAYPAPAPGVTPGPPSGPAGVPSQGPSAPGAYATGYPGAPGLTVPYNLP 121
DB 285 LPGLTG--NPGLPGRFGRFGFGFPVQGPFGPSPGPALEGFKNP--GQGPFGRPGLP 342
QY 122 GGVPFRL 129
DB 343 GPEGPRGL 350

RESULT 40
LEG7 MOUSE STANDARD; PRT; 135 AA.
ID LEG7_MOUSE
AC O54974;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-7 (Gal-7).
GN LGALS7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;

```



```

RX MEDLINE=98362649; PubMed=9697310;
RA Magnalido T., Fowles D., Darmon M.;
RT "Galectin-7, a marker of all types of stratified epithelia.";
RL Differentiation 63:159-168(1998).
CC
CC -1- FUNCTION: Could be involved in cell-cell and/or cell-matrix
CC interactions necessary for normal growth control. Pro-apoptotic
CC protein that functions intracellularly upstream of JNK activation
CC and cytochrome c release (By similarity).
CC
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear. May be secreted by
CC a non-classical secretory pathway (By similarity).
CC
CC -1- SIMILARITY: Belongs to the galectin (galapin/S-lectin) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: AF038562; AAB92566.1; -.
CC HSSP: P47929; 2CGL.
CC MGD: MGI:1316742; Lgal57.
CC InterPro: IPR008985; CONA_like_lect_57.
CC InterPro: IPR001079; Galectin.
CC Pfam: PF00337; Gal-bind_lectin; 1.
CC SMART: SM00276; GLCT; 1.
CC PROSITE: PS00309; GALACTIN; 1.
CC KEGG: Galectin; Lectin; Apoptosis; Nuclear protein.
CC INT MET 0 BY SIMILARITY.
CC FT BINDING 69 75 BETA-GALACTOSIDE (POTENTIAL).
CC FT SITE 135 AA; 15042 MW; 9C8E3CBF36C5013 CRC64;
CC SQ SEQUENCE
CC
Query Match 12.7%; Score 173; DB 1; Length 135;
Best Local Similarity 33.13; Pred. No. 0.00012;
Matches 42; Conservative 26; Mismatches 49; Indels 10; Gaps 5;
CC
QY 120 LPQGVVRMLITIGTVKPNANRIALDF---QRNDVAFHF--PRENNRRVIVCTXL 174
DB LPQGVVRMTWIRIGVLPDQGRFHVNLGSEBEGAAALHFNRLTSE---VFNTKQ 65
QY 175 DNKMGREBOSVFPFESGKPFKIOVLVDPHFYAV-NDAYLQYNHRVKKLNISKLGIS 233
DB 66 QGKMGREBREGTIGIPFORQGFVEVLITTEGFKAVVDDEYLHFNHRLPPA-RVRLVEVG 124
QY 234 GDIDLTS 240
DB 125 GDVQLHS 131
CC
RESULT 41
CA39 CHICK STANDARD; PRT; 675 AA.
ID CA39 CHICK
AC P32017;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3 (IX) chain precursor.
GN COL9A3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCB1_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92250566; PubMed=1577778;
RA Har-Bi R., Shatma Y.D., Aguilera A., Ueyama N., Wu J.J.,
RA Eyre D.R., Juricic L., Chandrasekaran S., Li M., Nah H.D.,
RA Upholt W.B., Tanzer M.L.;
RT "Cloning and developmental expression of the alpha 3 chain of chicken
RT type IX collagen.";

```

```

RL J. Biol. Chem. 267:10070-10076(1992).
CC
CC -1- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
CC COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(IX),
CC ALPHA 2(IX), AND ALPHA 3(IX).
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: M83179; AAB59960.1; -.
CC InterPro: IPR008161; C1g_helix.
CC InterPro: IPR008160; Collagen.
CC Pfam: PF01391; Collagen; 11.
CC ProDom: PD000007; C1g_helix; 3.
CC KEGG: Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal.
CC FT SIGNAL 1 21
CC FT CHAIN 22 675 COLLAGEN ALPHA 3(IX) CHAIN.
CC FT DOMAIN 25 515 TRIPLE-HELICAL REGION 3 (COL3).
CC FT DOMAIN 516 546 NONHELICAL REGION 3 (NC3).
CC FT DOMAIN 547 626 TRIPLE-HELICAL REGION 2 (COL2).
CC FT DOMAIN 627 631 NONHELICAL REGION 2 (NC2).
CC FT DOMAIN 632 658 TRIPLE-HELICAL REGION 1 (COL1).
CC FT DOMAIN 659 675 NONHELICAL REGION 1 (NC1).
CC FT SITE 242 244 CELL ATTACHMENT SITE (POTENTIAL).
CC FT SITE 591 593 CELL ATTACHMENT SITE (POTENTIAL).
CC FT CARBOHYD 479 479 N-LINKED (GLCNAC...) (POTENTIAL).
CC SQ SEQUENCE 675 AA; 63013 MW; C983FB9324A10098 CRC64;
CC
Query Match 12.7%; Score 173; DB 1; Length 675;
Best Local Similarity 32.24; Pred. No. 0.00052;
Matches 59; Conservative 14; Mismatches 64; Indels 46; Gaps 11;
CC
QY 9 DALSG-GNPNFGMGPGAMGN-QPAGAGYPGASYPGQAPGAYG---QAPGAY 62
DB 81 DGLTGDGPDPGPPGPDGRLGAPGPPGAGKGLPDPGPPGSGLPAGNGFGPPGP- 139
QY 63 HGAPGAYGAPAP-----GVY-----PGPPSGGAYPSSQGPAPG 98
DB 140 SGLPG-FPGPPGPPGAGLIPBGGDIQCPALCPGPPGPPGMPGFGHTGKGGPG 198
QY 99 AVA-TGPGYAPA--GELIVFYNLPLPGVVRMLITIGTVKPNANRIALDFORGNDAF 155
DB 199 EIGKEGKSGPPGPPGPIGVSGLG--PRGIRGLPGMPGAGR-----GDIGF 247
QY 156 HFP 158
DB 248 RGP 250
CC
RESULT 42
CA13 MOUSE STANDARD; PRT; 1464 AA.
ID CA13 MOUSE
AC P08121; Q61429; Q9C8N7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1 (III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6 X DBA; TISSUE=Embryo;

```

RX MEDLINE=95011609; PubMed=7926795;
 RA Toman D., de Crombrughe B.;
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and
 RT complete DNA sequence.";
 RL Gene 147:161-168(1994).
 [12]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6; TISSUE=Brain;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang Y., Hsieh P.,
 RA Dackiwko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stappleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Cantinci P., Prange C.,
 RA Raha S.S., Loquellon N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Roark S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.T., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodegryn E.J., Lu X., Gibbs S.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting R.M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.B., Jones S.U.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [13]
 RN SEQUENCE OF 1-488 FROM N.A.
 RP MEDLINE=8816785; PubMed=443309;
 RA Wood L., Theriault N., Vogeli G.;
 RT "Complete nucleotide sequence of the N-terminal domains of the murine
 RL alpha-1 type-III collagen chain.";
 RL Gene 61:225-230(1987).
 [14]
 RN SEQUENCE OF 1-28 FROM N.A.
 RP MEDLINE=85131189; PubMed=3972847;
 RA Liu G., Mudryj M., de Crombrughe B.;
 RT "Identification of the promoter and first exon of the mouse alpha 1
 RT (III) collagen gene.";
 RL J. Biol. Chem. 260:3773-3777(1985).
 [15]
 RN SEQUENCE OF 810-1464 FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RC MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Atawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Katsukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batelov S., Casavant T.,
 RA Pletschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bazh G.,
 RA Blake Y., Boffelli D., Bojunga N., Cantinci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guerinich S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Mombauts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [16]
 RN SEQUENCE OF 1442-1464 FROM N.A.
 RP STRAIN=C57BL/6;
 RC MEDLINE=91274355; PubMed=2054384;
 RA Metzenauer M., Roman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX

```

RT collagen mRNAs." ;
RL Blochim. Biophys. Acta 1089:241-243(1991) .
CC -I- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -I- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylysines.
CC -I- PM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains.
CC -I- PM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC the oxygen atom of a post-translationally added hydroxyl group (by
CC similarity). Contains 1 WFMC domain.
CC -I- SIMILARITY: Contains 1 WFMC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52046; CAA36279.1; -.
CC DR EMBL; BC043089; AAH34089.1; -.
CC DR EMBL; M858724; AAI58724.1; -.
CC DR EMBL; M18933; AAA37338.1; -.
CC DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
CC DR EMBL; AK019448; BAB31724.1; -.
CC DR EMBL; X57983; CAA41048.1; -.
CC DR PIR; A27353; A27353.
CC DR PIR; S59856; S59856.
CC DR MED; MG1:88453; Col3a1.
CC DR InterPro; IPRO08161; Ctg_helix.
CC DR InterPro; IPRO08160; Collagen.
CC DR InterPro; IPRO00885; Fbl_collagen_C.
CC DR InterPro; IPRO02181; Fibrinogen_C.
CC DR InterPro; IPRO01007; WFC_C.
CC DR Pfam; PF01410; COLFI; 1.
CC DR Pfam; PF01391; Collagen; 18.
CC DR ProDom; PD000007; Ctg_helix; 1.
CC DR ProDom; PD002078; Fbl_collagen_C; 1.
CC DR SMART; SMO0038; COLFI; 1.
CC DR SMART; SMO0214; WMC; 1.
CC DR PROSITE; PS01208; WFC_1; 1.
CC DR PROSITE; PS50184; WFC_2; 1.
CC KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 31 90 WFMC.
FT DOMAIN 155 169 NONHELIICAL REGION (N-TERMINAL).
FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
FT DOMAIN 170 1195 NONHELIICAL REGION (C-TERMINAL).
FT CARBOHD 262 262 O-LINKED (GAL. ...) (BY SIMILARITY).
FT MOD RES 262 262 HYDROXYLATION (BY SIMILARITY).
FT MOD RES 283 283 HYDROXYLATION (BY SIMILARITY).
FT MOD RES 859 859 HYDROXYLATION (BY SIMILARITY).
FT MOD RES 976 976 HYDROXYLATION (BY SIMILARITY).
FT MOD RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
FT MOD RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
FT DISULFD 1195 1195 INTERCHAIN (BY SIMILARITY).
FT DISULFD 1196 1196 INTERCHAIN (BY SIMILARITY).
SQ SEQUENCE 1464 AA; 13894 MW; 2104EC27A886090B CRC64;

Query Match      12.7%; Score 172.5; DB 1; Length 1464;
Best Local Similarity 32.5%; Pred. No. 0.0011;
Matches 62; Conservativity 7; Mismatches 61; Indels 61; Gaps 12;

```

```

Db 156 FDSYVKGSGVGMGTPGPBAGPPPG--PGSSSHPS--PGSPGYQGPGE-PGQAP 210
QY 60 GAYHAGPAGY-----PGAPAGVYPPPP--SGPGAY----- 89
Db 211 AGPPGPALGPAGAGKDGSGRGRORRORRCLPPPPGKPGAGMPGPMKMGHGFDR 270
QY 90 --SSGQPSAPGAYA---TGPYAPAA--GPLLVPYNLP.PGCVPRMTITLITGVKPNAN 141
Db 271 NGEKGTATGAPLKGNGLPDNGATGPMQPRGAPGERGRPG-----LEGAGARGN 321
QY 142 RIALDFRGND 152
Db 322 ---DGARGSD 328

RESULT 43
CC02 CAEEL
ID _CC02 CAEEL STANDARD; PRT; 301 AA.
AC P17656;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cuticle collagen 2 precursor.
GN COL-2 OR M01B6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=83050944; PubMed=7139711;
RA Kramer J.M., Cox G.N., Hirsch D.;
RT "Comparisons of the complete sequences of two collagen genes from
RT Caenorhabditis elegans.";
RL Cell 30:599-606 (1982).
EN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=85105075; PubMed=2578467;
RA Kramer J.M., Cox G.N., Hirsch D.;
RT "Expression of the Caenorhabditis elegans collagen genes col-1 and
RT col-2 is developmentally regulated.";
RL J. Biol. Chem. 260:1945-1951 (1985).
RN [3]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN=Bristol N2;
RA Kohara Y., Shin-I T., Suzuki Y., Sugano S., Thierry-Mieg D.,
RT "The Caenorhabditis elegans transcriptome project, a complementary
RT view of the genome.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Coles L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -!- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -!- TISSUE SPECIFICITY: Syncytial dorsal and ventral epidermis.
CC -!- DEVELOPMENTAL STAGE: Low level expression from embryos at comma
CC stage to adult; maximal level in later larvae.
CC -!- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01048; AAA27990.1; -.
DR EMBL; V00148; CAA23464.1; -.
DR EMBL; AY347575; AAP97869.1; -.
DR EMBL; Z68301; CAA92620.1; -.
DR PIR; B31219; B31219.
DR WormPep; M01B6.7; CE03759.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR Pfam; PF01391; Collagen_2.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL. 1 37
FT CHAIN 38 301 POTENTIAL.
FT DOMAIN 85 97 CUTICLE COLLAGEN 2.
FT DOMAIN 105 134 GLY-RICH.
FT DOMAIN 153 176 TRIPLE-HELICAL REGION.
FT DOMAIN 183 212 TRIPLE-HELICAL REGION.
FT DOMAIN 215 282 TRIPLE-HELICAL REGION.
SQ SEQUENCE 301 AA; 28025 MW; 33317E3BDAC302F9 CRC64;

Query Match 12.7%; Score 172; DB 1; Length 301;
Best Local Similarity 39.3%; Pred. No. 0.00029;
Matches 46; Conservative 3; Mismatches 60; Indels 8; Gaps 3;

QY 12 GSGGNPNPQGWPGAGWGNQPA---GAGYPGASYPGYPGAPPGAGPGAGPGAGYHAGPG 67
Db 180 SGPPGAPGPPGAPGAGNDGAPGAPGPGEPGASEGSGGEPGPPGPPGAPGAGNDGAPG 239
QY 68 AYPGAPAGVYPPGPPSGGPPGAPGAYATGPGAPAGPLIVPNLP.PGCV 124
Db 240 T--GGPFGPKKPGPAGAGAPGADONPGPG--TAGKFGGEGEKCTCKYCAIDGV 292

RESULT 44
SSA2 HUMAN
ID _SSA2 HUMAN STANDARD; PRT; 464 AA.
AC Q15428; O75245;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Splicing factor 3A subunit 2 (Spliceosome associated protein 62) (SAP
DE 62) (SF3A66).
GN SF3A2 OR SAP62.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94023929; PubMed=8211113;
RA Bennett M., Reed R.;
RT "Correspondence between a mammalian spliceosome component and an
RT essential yeast splicing factor.";
RL Science 262:105-108 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC lamerdin J.E., McCreedy P.M., Skowronski E., Adameon A.W.,
RC Burkhardt-Schultz K.J., Gordon L., Kyle A., Ramirez M., Seilwagen S.,
RC Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RC Krommiller B., Arellano A., Montgomey M., Carrano A.V.,
RC Liu S., Attix C., Andreise T., Trankheim M., Georgescu A., Avila J.,
RC Cofield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RC Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.5 kb contig in human 19p13.3 containing a
RT splice protease gene cluster";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

RC TISSUE=lung, and lymph;
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Mang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.D., Ustin J.B., Toshlyuk S., Carninci P., Prange C.,
 RA Bask S.A., Loquellano N.A., Peters G.J., Abramson R.D., Millar S.J.,
 RA Richard S., Morley K.C., McKernan K.J., Malek J.A., Gumarate P.H.,
 RA Villalón D.K., Muzny D.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP CHARACTERIZATION OF THE SPLICEOSOME.
 RX MEDLINE=20337962; PubMed=10882114;
 RA Das R., Zhou Z., Reed R.;
 RT "Functional association of U2 snRNP with the ATP-independent
 RT spliceosomal complex E";
 RL Mol. Cell 5:779-787 (2000).
 CC -I- FUNCTION: Subunit of the splicing factor SF3A required for 'A'
 CC complex assembly formed by the stable binding of U2 snRNP to the
 CC branchpoint sequence (BPS) in pre-mRNA. Sequence upstream to the
 CC binding of SF3A/SF3B complex upstream of the branch site is
 CC essential, it may anchor U2 snRNP to the pre-mRNA. May also be
 CC involved in the assembly of the 'E' complex.
 CC -I- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3A WHICH IS COMPOSED OF
 CC THREE SUBUNITS: SF3A3/SAP61, SF3A2/SAP62, SF3A1/SAP114. SF3A
 CC ASSOCIATES WITH THE SPLICING FACTOR SF3B AND A 12S RNA UNIT TO
 CC FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN COMPLEX (U2 snRNP).
 CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -I- SIMILARITY: BELONGS TO THE SF3A2 FAMILY.
 CC -I- SIMILARITY: Contains 1 matrix-type zinc finger.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L21990; AAA60301.1; -;
 DR EMBL: AC005263; AAC25613.1; -;
 DR EMBL: BC004434; AAH04434.1; -;
 DR EMBL: BC009903; AAH09903.1; -;
 DR PIR: A47655; A47655.
 DR Genew: HGNC:10766; SF3A2.
 DR CK: Q15428; -;
 DR MIM: 600796; -;
 DR GO: GO:0030532; C:small nuclear ribonucleoprotein complex; TAS.
 DR GO: GO:0005681; C:spliceosome complex; TAS.
 DR GO: GO:0006371; P:mRNA splicing; TAS.
 DR InterPro: IPR007087; Znf_C2H2.
 DR InterPro: IPR000690; Znf_matriin.
 DR InterPro: IPR003604; Znf_U1.
 DR SMART: SM00355; Znf_C2H2; 1.
 DR SMART: SM00451; Znf_U1; 1.
 DR PROSITE: PS00171; ZF_MATRIN; 1.
 KW Spliceosome; mRNA processing; mRNA splicing; Nuclear protein;
 KW Zinc-finger; Repeat.
 FT ZN_FTNG 54 84 MATRIN-TYPE.
 FT DOMAIN 233 236 POLY-PRO.
 FT DOMAIN 252 256 POLY-PRO.

FT DOMAIN 458 462 POLY-PRO.
 FT CONFLICT 29 29 R > P (IN REF. 1).
 SQ SEQUENCE 464 AA; 49255 MW; FA46F064A55E2CB CXC64;
 Query Match 12.6%; Score 171.5; DB 1; Length 464;
 Best Local Similarity 37.4%; Pred. No. 0.00046;
 Matches 49; Conservative 11; Mismatches 52; Indels 19; Gaps 7;
 QY 7 LHDALSSGNPNQCPQAGMGNQAPAGAGYGCAS--YGYGQAPPGAPPGAP--PGAY 62
 DB 294 VHPASGVHPAPGVHPAPGVHPAPGVHPPTSGVHPAPGVHPA--PGVHPAPGVH 351
 QY 63 HGAPGAYGAPAPGVYPCP-----PSGPAVPSG--QPSAPGAYATGPGAPACPLVY 115
 DB 352 PPAPGVHP--PAPGVHPSPSAGVHPQAGVHPAALVHPAPGVHPAPGVHPAPGVHP 409
 QY 116 YNLPDPGGVY 126
 DB 410 Q----PPGVHP 416
 RESULT 45
 CA24_HUMAN
 ID CA24_HUMAN STANDARD; PRT; 1712 AA.
 AC P08572;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 GN COL4A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8906769; PubMed=3198637;
 RA Hostalka S.L., Trygvason K.;
 RT "The complete primary structure of the alpha 2 chain of human type IV
 RT collagen and comparison with the alpha 1(IV) chain.";
 RL J. Biol. Chem. 263:19488-19493 (1988).
 RN [2]
 RP SEQUENCE OF 1-1042 FROM N.A.
 RC TISSUE=placenta;
 RX MEDLINE=88151998; PubMed=3345760;
 RA Brazel D., Poliner R., Oberhaeumer I., Kuehn K.;
 RT "Human basement membrane collagen (type IV). The amino acid sequence
 RT of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
 RT chain reveals deletions in the alpha 1(IV) chain.";
 RL Eur. J. Biochem. 172:35-42 (1988).
 RN [3]
 RP SEQUENCE OF 1254-1712 FROM N.A.
 RX MEDLINE=87219158; PubMed=3582677;
 RA Hostalka S.L., Kurkinen M., Trygvason K.;
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2
 RT chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
 RT (IV) but not with the collagenous domain or 3'-untranslated region.";
 RL FEBS Lett. 216:281-286 (1987).
 RN [4]
 RP SEQUENCE OF 1451-1485 FROM N.A.
 RX MEDLINE=87092438; PubMed=3025878;
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
 RT alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516 (1987).
 RN [5]
 RP SEQUENCE OF 1486-1712 FROM N.A.
 RX MEDLINE=87250571; PubMed=2449508;
 RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
 RT "Duplication of type IV collagen COOH-terminal repeats and species-
 RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
 RL J. Biol. Chem. 262:9231-9238 (1987).
 RN [6]

RP SEQUENCE OF 1-33 FROM N.A.
 RA MEDLINE=89034231; PubMed=3182844;
 RX Solinina R., Huotari M., Hostikka
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RT collagen are divergently encoded on opposite DNA strands and have an
 RT overlapping promoter region."
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [17]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89030632; PubMed=2846280;
 RA Poeschl E., Pollner R., Kuehn K.;
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
 RT basement membrane collagen type IV are arranged head-to-head and
 RT separated by a bidirectional promoter of unique structure."
 RL EMBO J. 7:2687-2695(1988).
 RN [8]
 RP SEQUENCE OF 1-33 FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=93305049; PubMed=8317999;
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
 RT "Identification of a novel sequence element in the common promoter
 RT region of human collagen type IV genes, involved in the regulation of
 RT divergent transcription."
 RL Biochem. J. 292:687-695(1993).
 RN [9]
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
 RC TISSUE=Placenta;
 RX MEDLINE=89005112; PubMed=2844531;
 RA Siebold B., Deutmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen."
 RL Eur. J. Biochem. 176:617-624(1988).
 CC -1- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 CC alpha 6(IV), each of which can form a triple helix structure
 CC with 2 other chains to generate type IV collagen network.
 CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the
 CC G-X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as there is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X05562; CAA29076.1; -;
 DR EMBL; X05610; CAA29098.1; -;
 DR EMBL; J02760; AAA58422.1; -;
 DR EMBL; M36963; AAA53099.1; -;
 DR EMBL; X12784; CAA31275.1; -;
 DR EMBL; J04217; AAA53097.1; -;
 DR PIR; A32024; CGH02B.
 DR Genew; HGNC:2203; COL4A2.
 DR MIM; 120090; -;
 DR GO; GO:0005587; C:collagen type IV; TAS.
 DR GO; GO:0005201; P:extracellular matrix structural constituent; TAS.
 DR GO; GO:000198; F:extracellular matrix organization and bioge. . .; NAS.
 DR InterPro; IPR008161; C1g_helix.

DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 24.
 DR ProDom; PD000007; C1g_helix; 7.
 DR ProDom; PD003923; Procollagnc4; 1.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Signal.
 FT SIGNAL 1 25
 FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
 FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.
 FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).
 FT DISULFD 1504 1593 OR 1590 (BY SIMILARITY).
 FT DISULFD 1537 1590 OR 1593 (BY SIMILARITY).
 FT DISULFD 1549 1555 BY SIMILARITY.
 FT DISULFD 1612 1708 OR 1705 (BY SIMILARITY).
 FT DISULFD 1646 1705 OR 1708 (BY SIMILARITY).
 FT DISULFD 1658 1665 BY SIMILARITY.
 FT CARBOHD 138 138 N-LINKED (GLCNAC. . .).
 FT CONFLICT 471 471 R -> P (IN REF. 2).
 FT CONFLICT 683 683 A -> G (IN REF. 2).
 FT CONFLICT 1575 1575 M -> I (IN REF. 5).
 FT CONFLICT 1663 1663 G -> H (IN REF. 9).
 FT CONFLICT 1701 1701 H -> G (IN REF. 9).
 SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
 Query Match 12.6%; Score 171.5; DB 1; Length 1712;
 Best Local Similarity 31.1%; Pred. No. 0.0015;
 Matches 51; Conservative 15; Mismatches 55; Indels 43; Gaps 7;
 QY 15 GNPNPDCMPGAWGNQPAAGAGYPGA-----GYP-----GYPGQ-----APP 50
 DB 1200 GLPPTGPPSGPSGDHEDPEPPPPGRRDPPGANTLPFGVGPQKGDGAPGERGPP 1259
 QY 51 GA-----YPGAPPPGAYHGAAPGAPGAPGV-----PGPPSG--PGAYPSSGQP 94
 DB 1260 GSPGLQGPPTGTPPSNINSAGC---DKGAPDIFPLKGYRGPPPGSALPLGSKDITNP 1316
 QY 95 SAPGAYATGPYGAPAGPLVYNTLPDGGVVRMLITLLGVKP 138
 DB 1317 GAPTPTGKMGAGDSGPGRGVPLGEXKPGRGEGMGNTGP 1360
 RESULT 46
 C413 HUMAN STANDARD; PRT; 1466 AA.
 AC P02461; Q15112;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kulvaneniemi H.,
 RA Prockop D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences."
 RL Biochem. J. 260:509-516(1989).
 RN [2]
 RP SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1

RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RN [3]
 RP SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 bromide peptides from the amino-terminal segment of type III collagen
 of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RN [4]
 RP REVISIONS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RN [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RN [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN [8]
 RP REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9]
 RP SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=609827;
 RA Lotiel H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RN [10]
 RP SEQUENCE OF 965-1200.
 RX MEDLINE=81208139; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 1(III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 RN [11]
 RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet M.J., Bernard M.P., Sippola M., Ramirez F.,
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 (III) collagen. Partial characterization of the 3' end region of the
 gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 RN [12]
 RP SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE=86187804; PubMed=3754462;
 RA Miskulin M., Dalgleish R., Kluever-Beckerman B., Remard S.I.,
 RT "Human type III collagen gene expression is coordinately modulated
 with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 RN [13]
 RP SEQUENCE OF 1-170 FROM N.A.

RC TISSUE=Placenta;
 RX MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 of human prepro alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RN [15]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X)
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [16]
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293988; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhatter S.L., Kleinert C.,
 RA Farley J.J., Zhang J., Noerregaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Rymaszewski M., Pearce W.H., Yao J.S.T.,
 RA Majumdar K., Smolens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.B., Michels V.V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 in the triple-helical domain of type III procollagen are an
 infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN [17]
 RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [18]
 RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [19]
 RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=94016385; PubMed=8411057;
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 661 by arginine in type III collagen
 produces mutant molecules with different thermal stabilities and
 causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 RN [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE=89109135; PubMed=2492273;
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 the alpha 1 (III) chain of type III procollagen exposes an arginine
 and causes Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RN [21]
 RP VARIANT EDS-IV VAL-960.
 RX MEDLINE=95268429; PubMed=7749417;
 RA Tromp G., de Paeppe A., Nuytink L., Madhatter S.L., Kuivaniemi H.,
 RT "Substitution of valine for glycine 793 in type III procollagen in
 Ehlers-Danlos syndrome type IV.";
 RL Hum. Mutat. 5:179-181(1995).
 RN [22]
 RP VARIANT EDS-IV GLU-1014.
 RX MEDLINE=92316511; PubMed=1352273;

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfaffler B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Baxu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Berns P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,
 Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemtson J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klip D., Lai Z.,
 Lako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yen R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 [2]
 REVISIONS.
 MEDLINE=22426069; PubMed=12537572;
 Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 Bertencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 Lewis S.E.;
 "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 systematic review.";
 Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 [3]
 SEQUENCE FROM N.A.
 STRAIN=Berkley; TISSUE=Head;
 MEDLINE=20196012; PubMed=10731138;
 Rubin G.M., Hong L., Brokstein P., Evans-Hoim M., Friese E.,
 Stapleton M., Harvey D.A.;
 "A *Drosophila* complementary DNA resource.";
 Science 287:2222-2224(2000).
 [4]
 SEQUENCE OF 77-90 AND 101-109, FUNCTION, SUBCELLULAR LOCATION, TISSUE
 SPECIFICITY, AND VARIANT.
 STRAIN=Canton-S;
 MEDLINE=21167515; PubMed=11267893;
 Lung O., Wolther M.P.;
 "Identification and characterization of the major *Drosophila*
melanogaster mating plug protein.";
 Insect Biochem. Mol. Biol. 31:543-551(2001).
 -1- FUNCTION: Major protein component of the posterior mating plug.
 Accessory gland proteins constitute, or are required for formation
 of the anterior mating plug. Posterior mating plug forms before
 sperm transfer and the anterior mating plug is formed after the
 start of mating.
 CC SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Specifically expressed in the ejaculatory bulb

and seminal fluid. Detected in mated females 3 minutes after the
 start of mating, and for at least 3 hours after the start of
 mating.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 DR EMBL: AE003466; AAF47319.2; -;
 DR EMBL: AF184225; AAD55736.1; -;
 DR FlyBase, FBgn0004181; Peb.
 KW Behavior; Signal; Polymorphism.
 FT SIGNAL 1 20
 FT CHAIN 21 377 POTENTIAL BULB SPECIFIC PROTEIN I.
 FT DOMAIN 116 256 EJACULATORY BULB SPECIFIC PROTEIN I.
 FT DOMAIN 159 377 GLY-RICH.
 FT VARIANT 84 G -> GANILAG (IN STRAIN CANTON-S).
 SO SEQUENCE 377 AA; 37819 MW; D020F027BD96B221 CRC64;
 Query Match 12.5%; Score 169.5; DB 1; Length 377;
 Best Local Similarity 34.9%; Pred. No. 0.0005;
 Matches 53; Conservative 10; Mismatches 46; Indels 43; Gaps 10;
 19 PQGNP-----GAMGNP-----AGAGVPGASYPGPGAPGAPGAPGAY 62
 129 PLGPMPLRPLPGPYPNRPPWLPINSPIRPGIFPGGPPSGPFGPS--PGSPSP-- 184
 63 HGAGAVPGAPAP-GVYPGPGSPGAYPSGQSPAPAVATG---PYGAPAPLIVPNYL 118
 185 ----GSPBGSGSPGSPBGSGS-PGG-PSPCGSPGPGFPBGSGSPSGGPFPGMQFPWIL 238
 119 -----PLPGGVPRMLITLITLGYKPN 140
 239 GGPSPNRGPRPFGILPGHLD--GSVPMS 267
 RESULT 49
 ANXB HUMAN STANDARD; PRT; 505 AA.
 ID ANXB HUMAN STANDARD; PRT; 505 AA.
 AC P50935;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Annexin A11 (Annexin XI) (calyculin-associated annexin 50) (CAP-50)
 DE (56 kDa autoantigen).
 GN ANXA11 OR ANX11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis; Carcinoma;
 RX MEDLINE=94140847; PubMed=7508441;
 RA Miskis Y., Pruijn G.J.M., van der Kemp A.W., van Venrooij W.J.;
 "The 56k autoantigen is identical to human annexin XI.";
 J. Biol. Chem. 269:4240-4246(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20469408; PubMed=11013079;
 RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
 Fernandez M.P.;
 "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
 RT annexins and source of orthologous cDNA isoforms.";
 RL Genomics 69:95-103(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bock S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Holik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Heaton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND POSSIBLY NUCLEAR.
 CC -1- DOMAIN: A pair of annexin repeats may form one binding site for
 CC calcium and phospholipid.
 CC -1- DISBASE: Antibodies against ANXA11 are present in sera from
 CC patients with various autoimmune diseases, predominantly in sera
 CC from patients with rheumatoid arthritis, systemic lupus
 CC erythematosus, or Sjogren's syndrome.
 CC -1- SIMILARITY: Belongs to the annexin family.
 CC -1- SIMILARITY: Contains 4 annexin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

DR EMBL; L19605; AAA19734.1; -;
 DR EMBL; AJ278463; CAB94995.1; -;
 DR EMBL; AJ278464; CAB94996.1; -;
 DR EMBL; AJ278465; CAB94997.1; -;
 DR EMBL; BC007564; AAH07564.1; -;
 DR PIR; A53152; A53152.
 DR HSSP; P13214; IACW.
 DR GENE; HGNC:535; ANXA11.
 DR MIM; 602572; -;
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0005635; C:nuclear membrane; NAS.
 DR GO; GO:0005654; C:nucleoplasm; NAS.
 DR GO; GO:0005543; F:phospholipid binding; TAS.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0006955; P:immune response; TAS.
 DR InterPro; IPR001464; Annexin.
 DR Pfam; PF00191; annexin; 4.
 DR PRINTS; PR00196; ANNEXIN; 4.
 DR ProDom; PD000143; Annexin; 4.
 DR SMART; SM00335; ANX; 4.
 DR PROSITE; PS00223; ANNEXIN; 4.
 KM Annexin; Calcium/phospholipid-binding; Repeat; Polymorphism.
 FT REPEAT 209 269 ANNEXIN 1.
 FT REPEAT 281 341 ANNEXIN 2.
 FT REPEAT 365 425 ANNEXIN 3.
 FT REPEAT 440 500 ANNEXIN 4.
 FT VARIANT 230 230 R -> C (in dbSNP:1049550).
 FT VARIANT 457 457 I -> V (in dbSNP:1802932).
 FT VARIANT 457 457 /FTID=VAR_012007.
 FT /FTID=VAR_012007.
 SQ SEQUENCE 505 AA; 54389 MW; 4ADCAC8F270BEE4 CRC64;

Query Match 12.5%; Score 169.5; DB 1; Length 505;
 Best Local Similarity 36.2%; Pred. No. 0.00066;
 Matches 55; Conservative 6; Mismatches 60; Indels 31; Gaps 7;

QY 11 LSGSGNPNPGQWPGAWGNQPAAGGVP--GASYPGYGQAPPG-----AYPGQ----- 56
 Db 1 MSYGVYPPPG-----GYPPAPAGGPGWGAAYPPPSMPPETIGDNVATYAGQNDYLS 55
 QY 57 ---APGAYHGA--PGATYGAPAPGYPPGPPSPGAYPSSGQPSAAGAYATGYGAPAG 110
 Db 56 GMAANSGTGGAGAMPNLYGAPGAPGYPVPVPPGFGQPPSAQGFVPP-----YGMVPPPG 111
 QY 111 --PLIVPYNPLPGGVYPRMLITLGVKQNA 140
 Db 112 NPESRMPSYPPYPCAPVPQGMPPPGQPPGA 143

RESULT 50
 CA14_HUMAN STANDARD; PRT; 1669 AA.
 AC P02462;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340433; PubMed=2701944;
 RA Souminen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RT type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [2]
 RP SEQUENCE OF 46-1257 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88083584; PubMed=3691802;
 RA Souminen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RT "Complete primary structure of the alpha 1-chain of human basement
 RT membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [3]
 RP SEQUENCE OF 1-943 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88029471; PubMed=3311751;
 RA Brazel D., Oberhauser I., Dieringer H., Babel W., Glanville R.W.,
 RA Deitzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 RT basement membrane collagen (type IV) reveals 21 non-triplet
 RT interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN [4]
 RP SEQUENCE OF 28-243.
 RX MEDLINE=86004708; PubMed=4043082;
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking
 RT region (75 domain) of the alpha 1 (IV) chain of human basement
 RT membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN [5]
 RP SEQUENCE OF 534-1447.
 RX MEDLINE=85003629; PubMed=6434307;
 RA Babel W., Glanville R.W.;
 RT "Structure of human-basement-membrane (type IV) collagen. Complete
 RT amino-acid sequence of a 914-residue-long pepsin fragment from the
 RT alpha 1(IV) chain.";
 RL Eur. J. Biochem. 143:545-556(1984).
 RN [6]
 RP SEQUENCE OF 1256-1669 FROM N.A.
 RX MEDLINE=85207819; PubMed=2581699;
 RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;

RT "cDNA clones coding for the pro-alpha(IV) chain of human type IV
 RT procollagen reveal an unusual homology of amino acid sequences in two
 RT halves of the carboxyl-terminal domain.";
 RL J. Biol. Chem. 260:7681-7687(1985).
 RN [7]
 RP SEQUENCE OF 1259-1669 FROM N.A.
 RX MEDLINE=85216555; PubMed=2582422;
 RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
 RA Kefauver N.A., Myers J.C.;
 RT "Restricted homology between human alpha 1 type IV and other
 RT procollagen chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89034231; PubMed=3182844;
 RA Soiminen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RT collagen are divergently encoded on opposite DNA strands and have an
 RT overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [9]
 RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
 RC TISSUE=Placenta;
 RX MEDLINE=89005112; PubMed=2844531;
 RA Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 RN [10]
 RP FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC [11]
 CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 CC alpha 6(IV), each of which can form a triple helix structure
 CC with 2 other chains to generate type IV collagen network.
 CC [12]
 CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the
 CC G-X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC [13]
 CC -1- PTM: Lysines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
 CC [14]
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC [15]
 CC -1- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC [16]
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL; M26576; AAA53098.1; JOINED.
 DR EMBL; J04217; AAA53098.1; JOINED.
 DR EMBL; M26550; AAA53098.1; JOINED.
 DR EMBL; M26540; AAA53098.1; JOINED.
 DR EMBL; M26542; AAA53098.1; JOINED.
 DR EMBL; M26543; AAA53098.1; JOINED.
 DR EMBL; M26544; AAA53098.1; JOINED.
 DR EMBL; M26545; AAA53098.1; JOINED.
 DR EMBL; M26546; AAA53098.1; JOINED.
 DR EMBL; M26547; AAA53098.1; JOINED.
 DR EMBL; M26537; AAA53098.1; JOINED.
 DR EMBL; M26538; AAA53098.1; JOINED.
 DR EMBL; M26548; AAA53098.1; JOINED.
 DR EMBL; M26549; AAA53098.1; JOINED.
 DR EMBL; M26551; AAA53098.1; JOINED.

DR EMBL; M26552; AAA53098.1; JOINED.
 DR EMBL; M26553; AAA53098.1; JOINED.
 DR EMBL; M26554; AAA53098.1; JOINED.
 DR EMBL; M26555; AAA53098.1; JOINED.
 DR EMBL; M26556; AAA53098.1; JOINED.
 DR EMBL; M26557; AAA53098.1; JOINED.
 DR EMBL; M26558; AAA53098.1; JOINED.
 DR EMBL; M26559; AAA53098.1; JOINED.
 DR EMBL; M26560; AAA53098.1; JOINED.
 DR EMBL; M26561; AAA53098.1; JOINED.
 DR EMBL; M26562; AAA53098.1; JOINED.
 DR EMBL; M26563; AAA53098.1; JOINED.
 DR EMBL; M26564; AAA53098.1; JOINED.
 DR EMBL; M26565; AAA53098.1; JOINED.
 DR EMBL; M26566; AAA53098.1; JOINED.
 DR EMBL; M26567; AAA53098.1; JOINED.
 DR EMBL; M26568; AAA53098.1; JOINED.
 DR EMBL; M26569; AAA53098.1; JOINED.
 DR EMBL; M26570; AAA53098.1; JOINED.
 DR EMBL; M26571; AAA53098.1; JOINED.
 DR EMBL; M26572; AAA53098.1; JOINED.
 DR EMBL; M26573; AAA53098.1; JOINED.
 DR EMBL; M26574; AAA53098.1; JOINED.
 DR EMBL; M26575; AAA53098.1; JOINED.
 DR EMBL; Y00706; CAA68698.1; -.
 DR EMBL; X05561; CAA29075.1; -.
 DR EMBL; M10940; AAA52006.1; -.
 DR EMBL; M11315; AAA52042.1; -.
 DR PIR; S16876; CGH04B.
 DR Genew; HGNC:2202; COL4A1.
 DR MIM; 120130; -.
 DR InterPro; IPR008161; C1g helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen_24.
 DR Pfam; PF000007; C1g helix; 6.
 DR ProDom; PD0003923; ProcollagenC4_1.
 DR SMART; SM00111; C4; 2.
 DR KX Extracellular matrix; Connective tissue; Basement membrane;
 DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172
 FT CHAIN 173 1669
 FT DOMAIN 173 1440
 FT DOMAIN 1441 1669
 FT CARBOHYD 126 126
 FT DISULFID 1460 1551
 FT DISULFID 1493 1548
 FT DISULFID 1505 1511
 FT DISULFID 1570 1665
 FT DISULFID 1604 1662
 FT DISULFID 1616 1662
 FT CONFLICT 237 238
 FT CONFLICT 241 241
 FT CONFLICT 319 241
 FT CONFLICT 719 719
 FT CONFLICT 837 837
 FT CONFLICT 842 842
 FT CONFLICT 896 896
 FT CONFLICT 904 904
 FT CONFLICT 914 914
 FT CONFLICT 998 998
 FT CONFLICT 1010 1010
 FT CONFLICT 1012 1012
 FT CONFLICT 1358 1358
 SQ SEQUENCE 1669 AA; 160611 MM; 3BBA6DFFB8B84 CRG4;
 Query Match 12.5%; Score 169; DB 1; Length 1669;
 Best Local Similarity 37.7%; Pred. No. 0.0021;

	Matches	55;	Conservative	9;	Mismatches	38;	Indels	44;	Gaps	11;
QY	14	SGNPNPQGM	PGAWGNQ	PAGAGGYPG	ASX-----	PGYPGQAPP	PGAYPGQAPP	PGAY	62	
Db	509	AGVPPDQCT	PGILG-QP	GAKGERGE	FFDLRLKGD	KDPPGFPGQ	-PG-MGR	-----	558	
QY	63	HGAPG-	AYGAPAPGV	YP-----	GPPSG--	PGAYPSSGQ	PSAPGAYAT	GPY--	105	
Db	559	-GSPGRDGH	PGILPGPKGS	PGSVGLK	GERGPPGV	GFPGRDTP	PPPPGYG	PPAGPI	GDK	617
QY	106	---	GAPAGPLT-	VPYNLPL	PGGVY	126				
Db	618	GQAGFP	GGGSPGL	PGPKGE	PGKIVP	643				

Search completed: August 23, 2004, 14:37:32
 Job time : 15 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:33:32 ; Search time 22 Seconds
(without alignments)
1075.596 Million cell updates/sec

Title: US-09-297-040-4
Perfect score: 1357
Sequence: 1 MADNPSLHDLATSGSGNPNPQ.....ISKLGISGIDLTLSASYTMI 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 65 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1315	96.9	250	2	A35820 galactin 3 - human
2	1082	79.7	264	2	A28651 galactose-specific
3	1082	79.7	264	2	A45983 lactose-binding le
4	1076.5	79.3	294	2	A49688 lactose-binding le
5	1075	79.2	262	2	A54889 IGF-binding protei
6	1049.5	77.3	245	2	A54909 carbohydrate-bind
7	1024	75.5	242	2	JC4300 galactin-3 - rabbi
8	437	32.2	139	2	S08576 lectin - mouse (fr
9	254	18.7	323	2	A55664 lectin L-36 - pig
10	250.5	18.5	145	2	A55932 galactin-5 - rat
11	236.5	17.4	324	2	A4631 lactose-binding le
12	223.5	16.5	317	2	JC6147 prostate carcinoma
13	218.5	16.1	316	2	A55975 galactin-8 - rat
14	211	15.5	136	2	I55469 galactin-7 - human
15	193	14.2	1049	1	CG8078 collagen alpha 1(I
16	192.5	14.0	463	2	S29170 annexin VII - mous
17	190.5	14.0	177	2	S65780 glycine/proline-ri
18	189.5	14.0	329	2	T32783 galactin-10 - hu
19	188.5	13.9	279	2	T37216 beta-galactoside-b
20	188.5	13.9	285	2	T26325 hypothetical prote
21	188	13.9	744	2	S15435 collagen alpha 1(I
22	187.5	13.8	313	2	T22828 hypothetical prote
23	186.5	13.7	179	2	A85217 hypothetical prote
24	186.5	13.7	277	2	T04441 amnextrin VII, long
25	186	13.7	488	1	LUMU7 collagen alpha 1(I
26	185.5	13.7	488	1	A27353 hypothetical prote
27	184	13.6	299	2	T19564 collagen alpha 1(I
28	183.5	13.5	743	1	S23779 hypothetical prote
29	183	13.5	299	2	T25407 hypothetical prote

ALIGNMENTS

RESULT 1	30	182.5	13.4	317	2	T29960	hypothetical prote
A35820	31	181.5	13.4	304	2	T26185	hypothetical prote
galactin 3 - human	32	181.5	13.4	304	2	T26185	hypothetical prote
N:Alernate names: CBP 35; epithelial-specific lectin 35; galactoside-binding lectin L-	33	180.5	13.3	356	2	T22827	hypothetical prote
C:Species: Homo sapiens (man)	34	180	13.3	744	1	A34246	collagen alpha 1(I
C>Date: 12-Oct-1990 #sequence revision 12-Oct-1990 #text_change 20-Apr-2000	35	180	13.3	744	1	S23298	collagen alpha 1(I
C:Accession: A35820; J00916; A47473; A36071; A45800	36	179	13.2	316	2	S08169	collagen col-12 pr
R:Robertson, M.W.; Albrandt, K.; Keller, D.; Liu, F.T.	37	179	13.2	316	2	S08170	collagen col-13 pr
Biochemistry 29, 8093-8100, 1990	38	178	13.1	1262	2	T22523	hypothetical prote
A>Title: Human IGF-binding protein: a soluble lectin exhibiting a highly conserved inte	39	177	13.0	299	2	T24833	hypothetical prote
A:Reference number: A35820; MUID:91084480; PMID:2261464	40	176.5	13.0	304	2	T22482	hypothetical prote
A:Accession: A35820	41	175.5	12.9	1357	2	T29265	hypothetical prote
A>Status: preliminary	42	175	12.9	569	2	S42886	collagen - sllkwr
A:Molecule type: mRNA	43	174.5	12.9	290	2	T24586	hypothetical prote
A:Residues: 1-250 <ROD>	44	174.5	12.9	296	2	A31219	collagen 1 - Caeno
A:Cross-references: GB:M57710; NID:g179530; PIDN:AAA5607.1; PID:g179531; GB:J02921	45	174.5	12.9	299	2	T29956	hypothetical prote
R:Oda, Y.; Leffler, H.; Sakakura, Y.; Kasai, K.; Baronides, S.H.	46	174	12.8	285	2	R88281	protein ZK892.1 li
Gene 99, 279-283, 1991	47	174	12.8	285	2	T29982	hypothetical prote
A>Title: Human breast carcinoma cDNA encoding a galactoside-binding lectin homologous t	48	173.5	12.8	266	2	T22706	hypothetical prote
A:Reference number: J00916; MUID:91216471; PMID:2022338	49	173.5	12.8	299	2	T22705	hypothetical prote
A:Accession: J00916	50	173.5	12.8	512	2	S70644	annexin VII - Afri
A:Molecule type: mRNA	51	173.5	12.8	754	2	A55267	collagen alpha 5(I
A:Residues: 1-250 <ODA>	52	172.5	12.7	1464	2	S59856	collagen alpha 1(I
A:Cross-references: GB:M56682; NID:g186921; PIDN:AAA36163.1; PID:g186922	53	172	12.7	283	2	T29980	hypothetical prote
R:Loetz, M.M.; Andrews Jr., C.W.; Korzeilus, C.A.; Lee, E.C.; Steele Jr., G.D.; Clarke,	54	172	12.7	301	2	T21314	hypothetical prote
Proc. Natl. Acad. Sci. U.S.A. 90, 3466-3470, 1993	55	172	12.7	301	2	B31219	collagen 2 - Caeno
A>Title: Decreased expression of Mac-2 (carbohydrate binding protein 35) and loss of it	56	172	12.7	675	2	S20819	collagen alpha 3(I
A:Reference number: A47473; MUID:93234518; PMID:7682704	57	171.5	12.6	464	2	A47655	spicosome-associ
A:Accession: A47473	58	171.5	12.6	1712	1	CGH02B	collagen alpha 2(I
A>Status: preliminary	59	171	12.6	290	2	T24590	hypothetical prote
A:Molecule type: mRNA	60	171	12.6	416	2	T32458	hypothetical prote
A:Residues: 1-250 <LOD>	61	170.5	12.6	314	2	T32985	hypothetical prote
A:Cross-references: GB:M57710; NID:g179530; PIDN:AAA5607.1; PID:g179531; GB:J02921	62	170	12.5	305	2	T20906	collagen alpha 2(I
R:Oda, Y.; Leffler, H.; Sakakura, Y.; Kasai, K.; Baronides, S.H.	63	170	12.5	469	2	A24450	collagen alpha 1(I
Gene 99, 279-283, 1991	64	170	12.5	1466	1	CGH07L	collagen alpha 1(I
A>Title: Human breast carcinoma cDNA encoding a galactoside-binding lectin homologous t	65	169.5	12.5	302	2	A31921	collagen dpy-13 pr
A:Reference number: A35820; MUID:91084480; PMID:2261464							

A/Cross-references: GB:S59012; NID:G299601; PIDN:AAB26229.1; PID:G299602
 A/Experimental source: normal colonic mucosa, colon carcinoma, cell line clone A
 A/Note: Sequence extracted from NCBI database (NCBI:129689, NCBI:129692)
 R/Cherayil, B.J.; Chaitovitz, S.; Wong, C.; Pillai, S.
 Proc. Natl. Acad. Sci. U.S.A. 87, 7324-7328, 1990
 A/Title: Molecular cloning of a human macrophage lectin specific for galactose.
 A/Reference number: A36071; PMID:90384999; PMID:2402511
 A/Accession: A36071
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-32; 'Q', 34, 'L', 37, 'RGFLSMGL', 46, 'RAGT', 51, 'R', 53-63, 'P', 65-87, 89-250 <CHE>
 A/Cross-references: GB:M53568; NID:G1196441
 A/Note: the sequence is revised in Genbank entry HUMMAC2, release 113.0, PIDN:AAA8086.1
 R/Raz, A.; Carmi, P.; Raz, T.; Hogan, V.; Mohamed, A.; Wolman, S.R.
 Cancer Res. 51, 2173-2178, 1991
 A/Title: Molecular cloning and chromosomal mapping of a human galactoside-binding protein
 A/Reference number: A49800; PMID:91183475; PMID:2009535
 A/Accession: A49800
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 1-32, 'Q', 34, 'LPGAGYPRAGTGLSTAPPGATMEVELRSTCTWSLTQ', 86-104, 'N', 106, 'W'
 A/Cross-references: GB:M64303; NID:G113862
 A/Note: this translation is not annotated in Genbank entry HUMGALBIN, release 113.0
 C/Genetics:
 A/Gene: GDB:LGALS3; MAC-2; LGALS2
 A/Cross-references: GDB:127515; OMIM:137033
 A/Map position: 1p13-1p13
 C/Superfamily: beta-galactoside-binding lectin
 C/Keywords: lectin; nucleus; phosphoprotein

Query Match 96.9%; Score 1315; DB 2; Length 250;
 Best Local Similarity 98.4%; Pred. No. 4, 4e-85;
 Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY	1	MADNFSLDALSGSGNPNQGMGAWGNQPAAGYPGASYPG-YPGQAPPGAYPGQAPP	59
DB	1	MADNFSLDALSGSGNPNQGMGAWGNQPAAGYPGASYPGAYPGQAPPGAYPGQAPP	60
QY	60	GAYHGAAPGAYPGAPAGYVPPSPSGGAYPSSGQAPAGYATGAYGAPAGLYTPYNL	118
DB	61	GAYHGAAPGAYPGAPAGYVPPSPSGGAYPSSGQAPAGYATGAYGAPAGLYTPYNL	120
QY	119	PLPGGVVPMILITLGTVCXNANRIALDFQRGNDVAFHF--PRENNRRRYVNTKLDNN	177
DB	121	PLPGGVVPMILITLGTVCXNANRIALDFQRGNDVAFHFENRRYVNTKLDNN	180
QY	178	WGRERQSVPPESGKPKYQVLVEPDHFKVANDAH-LQYHRYVKKLEISKLGISGDI	236
DB	181	WGRERQSVPPESGKPKYQVLVEPDHFKVANDAH-LQYHRYVKKLEISKLGISGDI	240
QY	237	DLTSASYTMI 246	
DB	241	DLTSASYTMI 250	

RESULT 2
 A28651
 galactose-specific lectin - mouse
 A/Alternate names: carbohydrate-binding protein 35; IGE-binding protein; lectin L-34; MA
 C/Species: Mus musculus (house mouse)
 C/Date: 28-Aug-1989 #sequence_revision 10-Feb-1995 #text_change 20-Aug-1999
 R/Cherayil, B.J.; Weiner, S.T.; Pillai, S.
 J. Exp. Med. 170, 1959-1972, 1989
 A/Title: The Mac-2 antigen is a galactose-specific lectin that binds IGE.
 A/Reference number: S08537; PMID:90063462; PMID:2584931
 A/Accession: S08537
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: mRNA
 A/Residues: 1-264 <CHE>
 A/Cross-references: EMBL:X16334; NID:G52986; PIDN:CAA34736.1; PID:G52987
 A/Note: this sequence was submitted to the EMBL Data Library, Oct-1989
 R/Jia, S.; Wang, J.L.

J. Biol. Chem. 263, 6009-6011, 1988
 A/Title: Carbohydrate binding protein 35. Complementary DNA sequence reveals homology wit
 A/Reference number: A28651; PMID:88198129; PMID:3360772
 A/Accession: A28651
 A/Molecule type: mRNA
 A/Residues: 'R', 3-264 <JIA>
 A/Cross-references: EMBL:J03723
 R/Raz, A.; Pazerini, G.; Carmi, P.
 Cancer Res. 49, 3489-3493, 1989
 A/Title: Identification of the metastasis-associated, galactoside-binding lectin as a ch
 A/Reference number: A37385; PMID:89275058; PMID:2525069
 A/Accession: A37385
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-3, 'T', 5-91, 'ST', 94-109, 'SAP', 113-264 <RAZ>
 A/Cross-references: GB:X16074; NID:G52850; PIDN:CAA34206.1; PID:G52851
 A/Note: authors translated the codon GAA for residue 219 as Ala, GAC for residue 221 as C
 R/Woo, H.J.; Shaw, L.M.; Messier, J.M.; Mercurio, A.M.
 J. Biol. Chem. 265, 7097-7099, 1990
 A/Title: The major non-integrin laminin binding protein of macrophages is identical to C
 A/Reference number: A35185; PMID:90236991; PMID:2332426
 A/Accession: A35185
 A/Molecule type: protein
 A/Residues: 159-163; 166-175; 214-226 <WOO>
 C/Superfamily: beta-galactoside-binding lectin
 C/Keywords: lectin; phosphoprotein

Query Match 79.7%; Score 1082; DB 2; Length 264;
 Best Local Similarity 78.0%; Pred. No. 9e-69;
 Matches 209; Conservative 16; Mismatches 17; Indels 26; Gaps 8;

QY	1	MADNFSLDALSGSGNPNQGMGAWGNQPAAGYPGASYPG-YPGQAPPGAYPGQAPP	59
DB	1	MADNFSLDALSGSGNPNQGMGAWGNQPAAGYPGASYPGAYPGQAPPGAYPGQAPP	59
QY	60	GAYHGAAPGAYPGAPAGYVPPSPSGGAYPSSGQAPAGYATGAYGAPAGLYTPYNL	100
DB	60	GAYHGAAPGAYPGAPAGYVPPSPSGGAYPSSGQAPAGYATGAYGAPAGLYTPYNL	116
QY	101	ANGPYGAPAPGPIIVPYNLPLPGGVVPMILITLGTVCXNANRIALDFQRGNDVAFHF-PR	159
DB	117	ANGPYGAPAPGPIIVPYNLPLPGGVVPMILITLGTVCXNANRIALDFQRGNDVAFHFENR	176
QY	160	FENNNRRYVNTKLDNNWGRERQSVPPESGKPKYQVLVEPDHFKVANDAH-LQYH	218
DB	177	FENNNRRYVNTKLDNNWGRERQSVPPESGKPKYQVLVEPDHFKVANDAH-LQYH	236
QY	219	HRVKKLEISKLGISGDI DLTSASYTMI 246	
DB	237	HRVKKLEISKLGISGDI DLTSASYTMI 264	

RESULT 3
 A45983
 lactose-binding lectin Mac-2 - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 03-May-1996
 R/Rosenberg, I.M.; Iyer, R.; Cherayil, B.; Chiodino, C.; Pillai, S.
 J. Biol. Chem. 268, 12393-12400, 1993
 A/Title: Structure of the murine Mac-2 gene. Splice variants encode proteins lacking fun
 A/Reference number: A45983; PMID:93286070; PMID:8509379
 A/Accession: A45983
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-264 <ROS>
 A/Cross-references: GB:L08649
 C/Genetics:
 A/Introns: 6/3; 128/3; 158/2; 209/2; 213/3
 C/Superfamily: beta-galactoside-binding lectin

Query Match 79.7%; Score 1082; DB 2; Length 264;
 Best Local Similarity 78.6%; Pred. No. 9e-69;

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

A;Accession: A55932
A;Status: preliminary
A;Molecule type: mRNA

A:Cross-references: GB:L36862; NID:g727175; PIDN:AAQ42050.1; PID:g727176
 R:Jung, S.K.; Fujimoto, D.
 J. Biochem. 116: 547-553, 1994
 A:Title: A novel beta-galactoside-binding lectin in adult rat kidney.
 A:Reference number: PX0077; MUID:95155264; PMID:7852273
 A:Accession: PX0077
 A:Molecule type: protein
 A:Residues: 7,'P',9-10,'T',12-19,'X',21-25;30-42;109-111,'N',113,'H',115,'VS',118-123,'K'
 A:Experimental source: kidney
 C:Comment: This protein exhibits activity to various saccharides and binds to Engelbreth
 C:Genetics:
 A:Gene: LGALS5
 C:Superfamily: beta-galactoside-binding lectin
 C:Keywords: acetylated amino end; lectin; monomer
 E:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental

Query Match 18.5%; Score 250.5; DB 2; Length 145;
 Best Local Similarity 40.4%; Pred. No. 6,4e-11;
 Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPNLPLPGGVPRMLITLGTVPKPNANRIALDFQGNVAFHF-PRFENNRR 166
 DB 9 PYPNLAPEPFSFENGKLPSTKIVISGVVLSDAKFPQNLKCGDIATFHLNPRFEN--- 65
 QY 167 VIVCNTKLDNNWGREERQ--SVFPESGKPKIQVLEPDHFKVAVNDAAH-LQYHRYKK 223
 DB 66 AVVRTQQLNNSWGPERSLPGSMFSPRGQRFVSWLICECHKVAVDQGHICEYSHRLMN 125

QY 224 LNEISKLGISGDIIDLT 239
 DB 126 LFDINTLEVAGDIQLT 141

RESULT 11
 A46631
 Lactose-binding lectin L-36 - rat
 N:Alternate names: galectin-4
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 29-Sep-1999
 C:Accession: A46631; S69096
 R:Oda, Y.; Herrmann, U.; Gilt, M.A.; Turck, C.W.; Burlingame, A.L.; Barondes, S.H.; Left
 J. Biol. Chem. 268, 5929-5939, 1993
 A:Title: Soluble lactose-binding lectin from rat intestine with two different carbohydrate
 A:Reference number: A46631; MUID:93194902; PMID:8449956
 A:Accession: A46631
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-324 <ODA>
 A:Cross-references: GB:M73553; NID:g294571; PIDN:AAA41505.1; PID:g294572
 A:Experimental source: intestine
 A:Note: sequence extracted from NCBI backbone (NCBIN:128409, NCBIPI:128410)
 R:Hardy, F.; Deviller, P.; Louisot, P.; Martin, A.
 FEBS Lett. 359, 169-172, 1995
 A:Title: Purification and characterization of the N-terminal domain of galectin-4 from r
 A:Reference number: S69096; MUID:95172227; PMID:7867792
 A:Accession: S69096
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 13-37;44-50,'E',52-66 <TR>
 C:Superfamily: lactose-binding lectin L-36
 C:Keywords: lectin

Query Match 17.4%; Score 236.5; DB 2; Length 324;
 Best Local Similarity 35.2%; Pred. No. 1,4e-09;
 Matches 68; Conservative 25; Mismatches 63; Indels 37; Gaps 10;

QY 71 GADPAGVVPQPSGPGAPYBSSGQPSAPGAVATGPYAP-----AGPLI---VPYNL 118
 DB 152 GGPAAAGYQGTWIP-AYBSAG-----YPPQMNSLPVPAGPPIENPYPYVG 198
 QY 119 PLPGGVPRMLITLGTVPKPNANRIALDFQGN--DVAHF-PRFENNRRVAVCNTKD 175
 DB 199 TLGGGLARRTLTIKGYVLPFAKNLITNFKVGSTGDIATFMRIGD---CVYRSTYNN 254

QY 176 NNWGREERQSV-PFESGKPKIQVLEPDHFKVAVNDAAH-LQYHRYKKLNEISKLGIS 233
 DB 255 GSGWSEERKLPYNPFGAGQFPDLSIRGTRFVKFVANGQHLDFSHRFOARFORVDMLEIK 314
 QY 234 GDIDLTASATYTKI 246
 DB 315 GDITL---SYVQI 324

RESULT 12
 JC6147
 prostate carcinoma tumor antigen 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 29-Sep-1999
 C:Accession: JC6147
 R:Su, Z.Z.; Lin, J.; Shen, R.; Fisher, P.B.; Goldstein, N.I.; Fisher, P.B.
 Proc. Natl. Acad. Sci. U.S.A. 93, 7252-7257, 1996
 A:Title: Surface-epitope masking and expression cloning identifies the human prostate c
 A:Reference number: JC6147; MUID:96293510; PMID:8692978
 A:Accession: JC6147
 A:Molecule type: mRNA
 A:Residues: 1-317 <SUA>
 A:Cross-references: GB:L78132; NID:g1932711; PIDN:AA851605.1; PID:g1932712
 C:Comment: This protein is a therapeutic reagent for intervention in pervasive and fata
 tein is a member of the galectin family.
 C:Superfamily: lactose-binding lectin L-36
 C:Keywords: tumor

Query Match 16.5%; Score 223.5; DB 2; Length 317;
 Best Local Similarity 39.3%; Pred. No. 1,1e-08;
 Matches 55; Conservative 23; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVPYNLPLPGGVPRMLITLGTVPKPNANRIALDFQGN-----DVAHF-PRFENNRR 165
 DB 16 VIFPFGTIPDQDPEPTLIVIRGHVPSDADRFQVDLQNGSSVXRADVAHFHNPFRFRAQ- 74
 QY 166 RYIVCNTKLDNNWGREERQSVFPESGKPKIQVLEPDHFKVAVNDAAH-LQYHRYKKL 224
 DB 75 -CIVCNTLTINERKGEETIYDPFPRKRSFELIVMLKDKQVAVANGKHTLLYGRIGP- 132

QY 225 NEISKLGISGDIIDLTASATY 244
 DB 133 EXIDTLTIGYKVNHSIGPS 152

RESULT 13
 A55975
 galectin-8 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 29-Sep-1999
 C:Accession: A55975
 R:Hardy, F.; Dekel, R.; Mestrovic, T.; Accili, D.; Zick, Y.
 J. Biol. Chem. 270, 3447-3453, 1995
 A:Title: Galectin-8, A new rat lectin, related to galectin-4.
 A:Reference number: A55975; MUID:9515445; PMID:7852431
 A:Accession: A55975
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-316 <HAD>
 A:Cross-references: GB:U09824; NID:g717031; PIDN:AAA66359.1; PID:g717032
 C:Superfamily: lactose-binding lectin L-36

Query Match 16.1%; Score 218.5; DB 2; Length 316;
 Best Local Similarity 39.9%; Pred. No. 2,5e-08;
 Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPPNPLPGGVPRMLITLGTVPKPNANRIALDFQGN-----DVAHF-PRFENNRR 166
 DB 16 IVPVSTITQOLKRGSLIVIRGHVPSDADRFQVDLQNGSLKPRADVAHFHNPFRKSN-- 73
 QY 167 VIVCNTKLDNNWGREERQSVFPESGKPKIQVLEPDHFKVAVNDAAH-LQYHRYKKLN 225
 DB 74 CIVCNTLTINERKGEETIYDPFPRKRSFELIVMLKKNFHVAVANGKHTLLYVHRINP-E 132

QY 226 EISKLGISGDIIDLTASV 243
 DB 133 KIDTLGIFKXNHSIGF 150

RESULT 14
 155469
 C:Species: Homo sapiens (man)
 C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 21-Jul-2000
 A:Accession: 155469
 R:Madson, P.; Rasmussen, H.H.; Flint, T.; Gromov, P.; Kruse, T.A.; Honore, B.; Vorum, H.
 J. Biol. Chem. 270, 5823-5829, 1995
 A>Title: Cloning, expression, and chromosome mapping of human galectin-7.
 A:Reference number: 155469; PMID:95197604; PMID:7534301
 A:Accession: 155469
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-136 <RES>
 A:Cross-references: GB:I07769; NID:g182131; PIDN:AAA67899.1; PID:g182132
 C:Genetics:
 A:Gene: GDB:LCAL57
 A:Cross-references: GDB:578907; OMIM:600615
 A:Map position: 19pter-19qter
 C:Superfamily: beta-galactoside-binding lectin

Query Match 15.5%; Score 211; DB 2; Length 136;
 Best Local Similarity 35.3%; Pred. No. 3.5e-08;
 Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;

QY 114 VPYNLPLPGGVVPRMLITITIGTVKPNANRTALDF---QKGNVAFHF-PRENNRRVY 168
 DB 4 VPKKSLPEGIRGTGTVLRIGLVPPNASHVNLGCEGSDAALHFNRLDTSE---V 60
 QY 169 VCNTKLDNNKRREROSVPPESGKPKIQVLEPDHFKAIVNDA-HLOYNHRVKKLNET 227
 DB 61 VENSKEQSGWGRERERPGVPFQRCQFEVLIIASDDQFAVVGDQYHFRRL-PLAVY 119
 QY 228 SKLGISGDIIDLT 240
 DB 120 RLVEVGSDVQLDS 132

RESULT 15
 CGB07S
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 07-May-1999
 A:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
 R:Fietzek, P.P.; Allmann, H.; Rautenberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
 A>Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen
 A:Reference number: A02862; PMID:80026026; PMID:488906
 A:Accession: A02862
 A:Molecule type: protein
 A:Residues: 1-242 <PIB>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
 A>Title: The covalent structure of calf skin type III collagen. II. The amino acid sequen
 A:Reference number: A38001; PMID:80026027; PMID:488907
 A:Accession: A38001
 A:Molecule type: protein
 A:Residues: 243-422 <DEW1>
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
 A>Title: The covalent structure of calf skin type III collagen. III. The amino acid sequen
 A:Reference number: A38002; PMID:80026028; PMID:488908
 A:Accession: A38002
 A:Molecule type: protein
 A:Residues: 423-571 <BN>
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979

A>Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequen
 A:Reference number: A38003; PMID:80026029; PMID:488909
 A:Accession: A38003
 A:Molecule type: protein
 A:Residues: 572-808 <LAN>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
 A>Title: The covalent structure of calf skin type III collagen. V. The amino acid sequen
 A:Reference number: A38004; PMID:80026030; PMID:488910
 A:Accession: A38004
 A:Molecule type: protein
 A:Residues: 809-947 <DEW2>
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
 A>Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequen
 A:Reference number: A38005; PMID:80026031; PMID:488911
 A:Accession: A38005
 A:Molecule type: protein
 A:Residues: 948-1049 <ALT>
 A:Experimental source: skin
 R:Henkel, W.
 Biochem. J. 318, 497-503, 1996
 A>Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
 A:Reference number: S71946; PMID:96404897; PMID:8809038
 A:Accession: S71946
 A:Molecule type: protein
 A:Residues: 87-106;1017-1029;1037-1049 <HEN>
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hy
 C:Superfamily: The type III collagen molecule is a trimer of identical chains, linked to each
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxyllysine; hydroxyprolin
 F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
 F:1-14/Region: amino-terminal nonhelical telopeptide
 F:15-1040/Region: helical
 F:587-589/Region: cell attachment (R-G-D) motif
 F:752-754/Region: cell attachment (R-G-D) motif
 F:775-777/Region: cell attachment (R-G-D) motif
 F:787-880/Region: cell attachment (R-G-D) motif
 F:935-937/Region: cell attachment (R-G-D) motif
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide
 F:95,107,119,938,950/Modified site: 5-hydroxyllysine (lys) #status experimental
 F:107,950/Modified site: allysine (lys) #status predicted
 F:107/Binding site: carbohydrate (lys) (covalent) #status experimental
 F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 14.2%; Score 193; DB 1; Length 1049;
 Best Local Similarity 33.7%; Pred. No. 5.5e-06;
 Matches 64; Conservative 8; Mismatches 58; Indels 60; Gaps 12;

QY 5 FEHLALSG-----SGNPNQGWPGAMGNQAGAGYRGAYRPG-QAPPAYPQA 57
 DB 2 YNAVIVKSSVAGGAGIAGVPGAPPPPP--PPGTSGHGA--PGAPGYQGPGR-PQA 56
 QY 58 PGAVHAGAPGAY-----PGAPAGVYRGP--SGPAPV----- 89
 DB 57 GPAGPPGPPGALGPGSKQSGRGRPRGRPPGPKMGKPGMGFGPKMKHGRGTGN 116
 QY 90 -SSGPPSAPGAY---ATGPGAPA--GPLVYNNLPLPGGVVPRMLITITIGTVKNNNR 142
 DB 117 GKSGEPGAPGALGNGVGDGAPGPMGRGAPGRGRPG-----LPGAAGARGN- 166
 QY 143 IALDFORGND 152
 DB 167 ---DEARSD 173

RESULT 16
 S29170
 A:Species: Mus musculus (house mouse)
 C:Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999
 A:Accession: S29170; S46209; S51173

R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
 Biochem. J. 289, 735-741, 1993
 A;Title: Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with ot
 A;Reference number: S29170; MUID:93168121; PMID:7916616
 A;Accession: S29170
 A;Molecule type: mRNA
 A;Residues: 1-463 <ZHA>
 A;Cross-references: EMBL:L13129
 R;Zhang-Keck, Z.Y.; Srivastava, M.; Kozak, C.A.; Caohuy, H.; Shirvan, A.; Burns, A.L.; F
 Biochem. J. 301, 835-845, 1994
 A;Title: Genomic organization and chromosomal localization of the mouse synexin gene.
 A;Reference number: S46209; MUID:94330961; PMID:8053909
 A;Accession: S46209
 A;Molecule type: DNA
 A;Residues: 1-463 <ZHF>
 R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
 submitted to the EMBL Data Library, May 1993
 A;Reference number: S51173
 A;Accession: S51173
 A;Molecule type: mRNA
 A;Residues: 1-144, 'S', 146-303, 'A', 305-463 <ZHW>
 A;Cross-references: EMBL:U13129; NID:9293293; PIND:AAA7238.1; PID:9293294
 C;Genetics:
 A;Gene: MGI:Anx7
 A;Cross-references: MGI:88031
 A;Map position: 14
 A;Intons: 19/3; 87/1; 121/1; 142/3; 177/1; 208/3; 246/3; 303/3; 360/3; 386/1; 423/3
 C;Superfamily: annexin VII; annexin repeat homology
 C;Keywords: calcium channel; ion channel
 F;163-234/Domain: annexin repeat homology <AX1>
 F;235-306/Domain: annexin repeat homology <AX2>
 F;318-390/Domain: annexin repeat homology <AX3>
 F;394-463/Domain: annexin repeat homology <AX4>

Query Match 14.2%; Score 192.5; DB 2; Length 463;
 Best Local Similarity 33.0%; Pred. No. 2.5e-06;
 Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

DB 11 LSGSNPNQGMGAMGNPAG-----AGGYPGASVYGGQAP---GAYPGAPG 60
 QY 1 MSYPPGP-PTGYPPPPGYPPACQESSFPTAGQTP-----YBSGPPMGAGAP-PABSG 52
 DB 61 AHGAPGAYP---GAPAGVYPG--PPSGGAYPS-----SGQPSAPG-AYA 101
 DB 53 GTPGA-GGYPAAGGYAPAGGYGALSPGAPPAVPGQGGAPGAGGSGYGPQPPQSYG 111
 QY 102 TGPY-----GAPAGPLVPPNLPFGVVP-----RMLITLGVKRNARKIADFOR 149
 DB 112 GGPAGVPPGPGFGGMPGQY---FGGAPYPSQPAAMTQGTGTILPASN---FDAMR 164
 QY 150 GNDVAFHPPFNMMNRVVCNTKUNMNGREERQSVFPESGK 193
 DB 165 DAEILRRKAMKFGGTDEQALVDVVSNSRNDQROQIKAFKTMGK 208

RESULT 17
 S65780
 glycine/proline-rich protein GPR - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
 C;Accession: S65780
 R;Maruy, I.; Montfort, A.; Stiefel, V.; Ludevid, D.; Delserny, M.; Puigdomenech, P.
 Plant Mol. Biol. 30, 625-636, 1996
 A;Title: Molecular characterization of the gene coding for GPR, a class of proteins ric
 A;Reference number: S65780; MUID:9619273; PMID:8605310
 A;Accession: S65780
 A;Molecule type: DNA
 A;Residues: 1-177 <MAR>
 A;Cross-references: EMBL:X84315; NID:9886425; PIND:CAAS9059.1; PID:91465364
 A;Note: mRNA was also sequenced
 C;Genetics:
 A;Gene: GPRP
 A;Intons: 99/1

Query Match 14.0%; Score 190.5; DB 2; Length 177;
 Best Local Similarity 44.6%; Pred. No. 1.2e-06;
 Matches 50; Conservative 4; Mismatches 41; Indels 17; Gaps 6;

DB 13 GSGNPNQGMGAMGNPAG---AGYPPGASYP---GPGQAPRGAAPGQA--PPGAYHGA 66
 QY 2 GGDNDNDK-KGFNGYPPAGYPPGAYPPAGYPPQGGYP--PPGAYPPAGYPPGAYPPAP 58
 DB 67 GAYPAPAPGYPPGPPS-----GPGAYPSGQPSAPGAYATGYPGAPAG 110
 QY 59 GGYPPAPRGYGPAPRGYGPAPRGYGPAPRGYGPAPRGYGPAPRGYGPAPRGYGPAPRGY 110

RESULT 18
 T32783
 hypothetical protein C50D2.4 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 15-Sep-2003
 C;Accession: T32783
 R;Sammons, L.; Woldmann, P.; Bauer, C.
 submitted to the EMBL Data Library, December 1997
 A;Description: The sequence of C. elegans coemid C50D2.
 A;Reference number: Z21224
 A;Accession: T32783
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-329 <SAM>
 A;Cross-references: EMBL:AF040642; PIND:AA94952.1; GSPDB:GN00020; CESP:C50D2.4
 A;Experimental source: strain Bristol N2; clone C50D2
 C;Genetics:
 A;Gene: CESP:C50D2.4
 A;Map position: 2
 A;Intons: 27/1; 59/3; 76/3; 237/1

Query Match 14.0%; Score 189.5; DB 2; Length 329;
 Best Local Similarity 43.4%; Pred. No. 2.8e-06;
 Matches 53; Conservative 1; Mismatches 45; Indels 23; Gaps 7;

DB 19 POGWPGAMNQ-PAGAGGYGASVYGP-----QAPPGAYPPQAPPGA-----YHGAPGA 68
 DB 187 PTPGPPGPPKQGPKNPNERGA--POTGGMALPPGPBPBPQGGPAGDGLPQNPAGAPV 244
 QY 69 -----YGPAPAGVYPPPSGPGAYPSGQPSAPGAYATGYPGAPAGLI--VPYNLPL 120
 DB 245 PGQVIEVPTPAPA---GPPBPBPBGPGGPGAGGSSQPGGPGPDGIDGAPNPSS 301
 QY 121 PG 122
 DB 302 PG 303

RESULT 19
 T37216
 beta-galactoside-binding protein GBP - Caenorhabditis elegans
 N;Alternate names: beta-galactoside-binding lectin homolog, 32K
 C;Species: Caenorhabditis elegans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
 C;Accession: T37216; T26324; A42846; S27798
 J. Biol. Chem. 272, 26669-26677, 1997
 R;Arata, Y.; Hirabayashi, U.; Kasai, K.
 A;Title: Structure of the 32-kDa galactin gene of the nematode Caenorhabditis elegans.
 A;Reference number: Z21635; MUID:97476274; PMID:9334250
 A;Accession: T37216
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-279 <ARA>
 A;Cross-references: EMBL:AB000802; NID:92564035; PIND:BA22942.1; PID:92564036
 R;Smey, R.
 submitted to the EMBL Data Library, November 1996
 A;Reference number: Z20198
 A;Accession: T26324
 A;Status: preliminary; translated from GB/EMBL/DDBJ

Query Match	13.9%	Score 188.5;	DB 2;	Length 279;
Best Local Similarity	36.4%	Pred. No. 2.8e-06;		
Matches	47;	Conservative	26;	Mismatches 49;
			Indels	7;
			Gaps	5;
QY	114	VPYMLPFGVAVPRMLITILGTVEKNARIALDFQRGN-DVAFFHP-PRKNENRRVIVYCN	171	
		: : : : : : : : : : : : : : : : : : : :		
Db	150	VFPESGLANGLPVQKSLVFGVEKKARFHYNLIRKKGDSISFHHNPFDEGH--VIRN	206	
		: : : : : : : : : : : : : : : : : : : :		
QY	172	TKLDNNWGREEROSVFPPESGKPFKIQLVBEEDHKFAVN-DAILQYHHRVYKLNIEIKL	230	
		: : : : : : : : : : : : : : : : : : : :		
Db	207	SLANWEMNEBEREGKNPFPEKVGDFGLVIONENYAFQVFNGBRIRYSFHRADP-HDIAGL	265	
		: : : : : : : : : : : : : : : : : : : :		
QY	231	GISGDIIDL T 239		
		: : : :		
Db	266	QISGDIELS 274		
		: : : :		

Query Match	13.9%	Score 188.5	DB 2	Length 285
Best Local Similarity	36.4%	Pred. No. 2.8e+06		
Matches 47, Conservative	26	Mismatches 49	Indels 7	Gaps 5

QY	231	GISGDIDL	239
		:	:
Db	272	QISGDIELS	280

Query Match	13.9%	Score 188	DB 2	Length 744	
Best Local Similarity	30.5%	Pred. No. 8.6e-06			
Matches	60	Conservative	12	Mismatches	67
				Indels	58
				Gaps	9
QY	15	GNPNPGCMPGAMG----	NQAPAGAGYGC-----	ASYGYPG-----	46
Db	429	GEFGLQGFPPKRGFLTEGVGPPGMKGFPPGLIGPKGKHGKGVPLPGVFGLLGPKKEPGP			488
QY	47	-----QAPRAYPQQAPPPGAVHAGP----	AYGAPAPGYP-----	GPBGP	85
Db	489	GDQGLQGPFG--TPGIIGPSGPIGPBGLRPGRGEGELGPPPPFGIGKGVAGLHGPBPKP			547
QY	86	GAYESSGQSPAGAYATGPGAPGLIVPNLPLPGVAVPRMLITLIGTAKPNRATL			145
Db	548	GALGPQGGPGLPG--PPGPDPGPSPRAVMPPTPPQGPBYLPLDKGLGIDGVAFPHAT-----			601
QY	146	DFQRGND--VAHFHPPPT	160		
Db	602	GAKKGKNKGPAYPEMPAF	618		

Query Match	13.8%;	Score 187.5;	DB 2;	Length 313;
Best Local Similarity	43.9%;	Pred. No. 3.7e-06;		
Matches 50; Conservative	2;	Mismatches 41;	Indels 21;	Gaps 6;

QY 14 SGNPNPGMGWGNQAGAGYR----GASVPGPGQAPPGAGVPGQAPPGAYHAGAPG- 68
 DB 172 AGPPGPGAPGQKQ--PSGAPGVGQGTGASLPPEPPAPGPPGSGGPGSGNGAGAPGAP 229
 QY 69 -----YCGAPAPGVYPPGPPSGPGAYP-----SSGQPSAPGAVATGPGAPAGP 111
 DB 230 GQYVDVGTGTGAPAPPPPP- GPAGAPQPGQSSGQGGGCG--PQGNAGAPGAP 280

RESULT 23

A85217
 hypothetical protein AT4g19200 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
 C:Accession: A85217
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
 A:Reference number: A85001; MUID:20083488; PMID:10617158
 A:Accession: A85217
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-179 <STD>
 A:Cross-references: GB:NC_001268; NID:97268715; PIDN:CAB76922.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: AT4g19200
 A:Map position: 4

Query Match 13.7%; Score 186.5; DB 2; Length 179;
 Best Local Similarity 45.7%; Pred. No. 2.4e-06;
 Matches 43; Conservative 5; Mismatches 31; Indels 15; Gaps 5;

QY 21 GMPGAMGNQAPGAGYPGASYPGYPGQAPPGAPPGQA-PPGAYHAGAPPGAPPGYYP 79
 DB 18 GPPGG-GHYPPAGGYPGQYPPQGGYPPAGGYPGAGYPPGAYPPGAPPGAYP--PAPGGY 74
 QY 80 GPPSGGAYPPSSGQPSAPGAVATGPGYAPAGPLI 113
 DB 75 -----PAGY-----APGAHSHSGSGGLGAMI 97

RESULT 24

T04441
 hypothetical protein T18B16.170 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T04441
 R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: Z15359
 A:Accession: T04441
 A:Molecule type: DNA
 A:Residues: 1-277 <BEV>
 A:Cross-references: EMBL:AL021687
 A:Experimental source: cultivar Columbia; BAC clone T18B16
 C:Genetics:
 A:Map position: 4
 A:Introns: 87/1; 144/1; 179/1; 207/3
 A:Note: T18B16.170

Query Match 13.7%; Score 186.5; DB 2; Length 277;
 Best Local Similarity 45.7%; Pred. No. 3.8e-06;
 Matches 43; Conservative 5; Mismatches 31; Indels 15; Gaps 5;

QY 21 GMPGAMGNQAPGAGYPGASYPGYPGQAPPGAPPGQA-PPGAYHAGAPPGAPPGYYP 79
 DB 18 GPPGG-GHYPPAGGYPGQYPPQGGYPPAGGYPGAGYPPGAYPPGAPPGAYP--PAPGGY 74
 QY 80 GPPSGGAYPPSSGQPSAPGAVATGPGYAPAGPLI 113
 DB 75 -----PAGY-----APGAHSHSGSGGLGAMI 97

RESULT 25

LNUH7
 annexin VII, long form - human
 N:Alternate names: synexin
 N:Contents: annexin VII, long form; annexin VII, short form
 C:Species: Homo sapiens (man)
 C>Date: 30-Sep-1992 #sequence_revision 26-Jan-1996 #text_change 22-Jun-1999
 C:Accession: A54467; A32554; A39513; B39513
 R:Shirvan, A.; Srivastava, M.; Wang, M.G.; Cultraro, C.; Magendzo, K.; McBride, O.W.; P.
 Biochemistry 33, 6888-6901, 1994
 A:Title: Divergent structure of the human synexin (annexin VII) gene and assignment to
 A:Reference number: A54467; MUID:94264005; PMID:7515686
 A:Accession: A54467
 A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA
 A:Residues: 1-488 <SHI>
 R:Burns, A.L.; Magendzo, K.; Shirvan, A.; Srivastava, M.; Rojas, E.; Alijani, M.R.; Pol
 Proc. Natl. Acad. Sci. U.S.A. 86, 3798-3802, 1989
 A:Title: Calcium channel activity of purified human synexin and structure of the human
 A:Reference number: A32554; MUID:89264510; PMID:2542947
 A:Accession: A32554
 A:Molecule type: mRNA
 A:Residues: 1-145,168-488 <BUR>
 A:Cross-references: EMBL:J0543; NID:9338243; PIDN:AAA36616.1; PID:9338244
 R:Magendzo, K.; Shirvan, A.; Cultraro, C.; Srivastava, M.; Pollard, H.B.; Burns, A.L.
 J. Biol. Chem. 266, 3228-3232, 1991
 A:Title: Alternative splicing of human synexin mRNA in brain, cardiac, and skeletal mus
 A:Reference number: A39513; MUID:91131630; PMID:1825209
 A:Accession: A39513
 A:Molecule type: mRNA
 A:Residues: 137-145,168-176 <MAG>
 A:Cross-references: EMBL:J05732

A:Accession: B39513
 A:Molecule type: mRNA
 A:Residues: 137-176 <MA2>
 C:Comment: Annexin undergo reversible, calcium-dependent binding to membrane phospholi
 derolipid.
 C:Comment: The long form of annexin VII is more prevalent in brain, heart, and skeletal

C:Genetics:
 A:Gene: GDB:ANX7
 A:Cross-references: GDB:369042; OMIM:186360
 A:Map position: 10q21.1-10q21.2
 A:Introns: 17/3; 87/1; 124/1; 145/3; 167/3; 202/2; 233/3; 271/3; 328/3; 385/3; 411/1; 4
 C:Superfamily: annexin VII; annexin repeat homology
 C:Keywords: alternative splicing; calcium binding; calcium channel; duplication; endome
 F:1-488/Product: annexin VII, long form #status predicted <MA1>
 F:1-145,168-488/Product: annexin VII, short form #status predicted <MA3>
 F:1188-259/Domain: annexin repeat homology <AX1>
 F:1199-215/Region: endonexin fold #status predicted
 F:260-331/Domain: annexin repeat homology <AX2>
 F:271-287/Region: endonexin fold #status predicted
 F:343-415/Domain: annexin repeat homology <AX3>
 F:355-371/Region: endonexin fold #status predicted
 F:419-488/Domain: annexin repeat homology <AX4>
 F:430-446/Region: endonexin fold #status predicted

Query Match 13.7%; Score 186; DB 1; Length 488;
 Best Local Similarity 41.4%; Pred. No. 7.5e-06;
 Matches 48; Conservative 2; Mismatches 44; Indels 22; Gaps 6;

QY 17 PNPQGMGAMGNQAPGAGYPGASYPGYPGQAPPGAPPGQA-PPGAYHAGAPPGAPPG 76
 DB 33 PVPSSGPP-----PMGAGAYPPVPPSSGYPG---AGGYPA---PGGY-PAPGGYPPGAPPG 79
 QY 77 VYPPSSGPPGAYPPSSGQPSAPGAVATGPGYAP-----AGPLVPYNLPLPGGVP 126
 DB 80 ---GASVPGVPGGGFGVPPGAGGSGYPPQPPSGSYGGGAPVPLPGRPPGGMP 132

RESULT 26

A27353
 collagen alpha 1(III) chain precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)
C>Date: 28-Aug-1999 #sequence_revision 30-Sep-1990 #text_change 13-Aug-1999
C/Accession: A27353; A22287; P50066
R/Mood, L.; Theriault, N.; Vogeli, G.
Gene 61, 225-230, 1987
A>Title: Complete nucleotide sequence of the N-terminal domains of the murine alpha-1(I)
A/Reference number: A91588; MUID:86167858; PMID:3443309
A/Accession: A27353
A/Molecule type: mRNA
A/Residues: 1-488 <MOO>
A/Cross-references: GB:M18933; NID:9192276; PIDN:AAA37338.1; PID:g553883
A/Note: in the helical region, Gly-X-Y repeat is interrupted by 287-Val instead of Gly
R/Liau, G.; Mody, M.; de Crombrugne, B.
J. Biol. Chem. 260, 3773-3777, 1985
A>Title: Identification of the promoter and first exon of the mouse alpha 1(III) collage
A/Reference number: A92513; MUID:85131189; PMID:3972847
A/Accession: A22287
A/Molecule type: DNA
A/Residues: 1-28 <LTA>
C/Genetics: 29/1
A/Intons: 29/1
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C/Keywords: coiled coil; extracellular matrix; glycoprotein; sulfoprotein; trimer; trip
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-154/Domain: amino-terminal propeptide #status predicted <APR>
F:155-488/Product: collagen alpha 1(III) chain C repeat homology <VWC>
F:164-488/Region: helical (fragment)
F:40,152,159/Binding site: sulfate (Tyr) (covalent) #status predicted

Query Match 13.7%; Score 185.5; DB 2; Length 488;
Best Local Similarity 36.5%; Pred. No. 8.2e-06;
Matches 57; Conservative 10; Mismatches 46; Indels 43; Gaps 10;

QY 5 FSLHDLISG-----SGPNPQGMFGAMGNPAGAGVPSGYSYRPS-QAPPGAYPGQAP 59
DB 156 FDSYDVKSQVGMAGVPGAGPPGLPG--PSSSGHPS--PSSPYQGPFGP-PQQAQP 210
QY 60 GAYHGAPGAY-----PGAPAPGVYPPGP--SGPGAYSSGQSPAPGAYA-- 101
DB 211 AGPPPGPAGLPGAPGAGKDGSGRPGRPGRGLPGPPGKIGP-----AGMPGPPGKGR 265
QY 102 -----TGPYAPR--GELIVPYNLPLPGGVNPR 127
DB 266 GFDGRNGEKETGAPGLKGENVLPQDNGAPGPMGPR 301

RESULT 27

T19564
Hypothetical protein C29F4.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T19564
R/Kershaw, J.
Submitted to the EMBL Data Library, January 1996
A/Reference number: Z19143
A/Accession: T19564
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-299 <WIL>
A/Cross-references: EMBL:Z68335; PIDN:CAA92729.1; GSPDB:GN00022; CESP:C29F4.1
A/Experimental source: clone C29F4
C/Genetics:
A/Gene: CESP:C29F4.1
A/Map position: 4
A/Intons: 27/3

Query Match 13.6%; Score 184; DB 2; Length 299;
Best Local Similarity 42.9%; Pred. No. 6.2e-06;
Matches 54; Conservative 4; Mismatches 46; Indels 22; Gaps 8;

QY 15 GNPNDQGMFGAMGNPAGAGVPSGYSYRPS-QAPPGAYPGQAP 66

DB 172 GARGDQPPAPKKGSGNPNBAPGAPGQPGQDAPSPPIPPGA-PGAPGAPGQGP 230
QY 67 GA-----YGPAPGVYPPPPSGGAYSSGQSPAPGAYATGYPGAPGPIVYNL 118
DB 231 GARGPQHDGQPPAPGP-----KGNPNBAPGAGDGNPAPGAPGAGGAGKGG-ICPKYC 265
QY 119 PLPGGV 124
DB 286 ALDGGV 291

RESULT 28

S23779
collagen alpha 1(VIII) chain - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: S23779
R/Muragaki, Y.; Shiota, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Ninomiya, Y.
Eur. J. Biochem. 207, 895-902, 1992
A>Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypepti
A/Reference number: S23779; MUID:92362626; PMID:1499564
A/Accession: S23779
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-743 <WUR>
A/Cross-references: EMBL:X66976; NID:g50493; PIDN:CAA47387.1; PID:g1359953
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
F:16-742/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 13.5%; Score 183.5; DB 1; Length 743;
Best Local Similarity 34.6%; Pred. No. 1.8e-05;
Matches 55; Conservative 8; Mismatches 57; Indels 39; Gaps 8;

QY 15 GNPNDQGMFGAMGNPAGAGVPSGYSYRPS-QAPPGAYPGQAP 57
DB 447 GPPGMRGLPPIG--PKGGHKGHLGLGVYGLGPKGEPGLPDQGLQGPFG-IPGLV 503
QY 58 PPGAYHGAPG-----AYPAPAPGVY-----GPPSGPAGYSSGQSPAPGAYA 101
DB 504 GPGSPGPPGIPGPGPGLPGLPGPFGVGRKPVAGLHGPGRGALGPQGGJFG--P 561
QY 102 TGPYAPR--GELIVPYNLPLPGGVNPRMLITLIGTKFPA 140
DB 562 PGPDPGPPAPVMPPIPS--QGEYLPDMGLGIDGVTPIPA 599

RESULT 29

T25407
Hypothetical protein T28C6.4 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T25407
R/Lloyd, C.
Submitted to the EMBL Data Library, September 1995
A/Reference number: Z20030
A/Accession: T25407
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-299 <WIL>
A/Cross-references: EMBL:Z54238; PIDN:CAA90995.1; GSPDB:GN00022; CESP:T28C6.4
A/Experimental source: clone T28C6
A/Accession: T25409
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-299 <WIL>
A/Cross-references: EMBL:Z54238; PIDN:CAA90995.1; GSPDB:GN00022; CESP:T28C6.6
C/Genetics:
A/Gene: CESP:T28C6.4; CESP:T28C6.6
A/Map position: 4
A/Intons: 27/3

Query Match 13.5%; Score 183; DB 2; Length 299;

Best Local Similarity 42.9%; Pred. No. 7.3e-06;
Matches 54; Conservative 4; Mismatches 46; Indels 22; Gaps 8;
QY 15 GNPNGQGMFGAMGNO-PAGAGGYPGA-SYGYPGQ-AP-----PGAYPGQAPPGAYHGA 66
DB 172 GAPGNDGPGAGKPKSPSPNPGAPGAPGQPGDAPSEPIITPGA-PGAPGAPGQGP 230
QY 67 GA-----YGPAPAGVYPGPPSGPGAYPSSGQSPAGATGPGYGAAPGLIYPNLP 118
DB 231 GAPGCGHGDGQGPAPG---KGPNGNPGAPGADGNPGAGQAGTSGAGEKG--ICPKYC 285
QY 119 PLPGV 124
DB 286 AIDGV 291

RESULT 30
T23960
hypothetical protein F36A4.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T23960
R/Miller, N.; Bradshaw, H.
submitted to the EMBL Data Library, April 1996
A/Description: The sequence of C. elegans cosmid F36A4.
A/Reference number: Z20713
A/Accession: T23960
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-317 <MIL>
A/Cross-references: EMBL:U53333; PIDN:AAA96159.1; GSPDB:GN00022; CESP:F36A4.6
A/Experimental source: strain Bristol N2; clone F36A4
C/Genetics:
A/Gene: CESP:F36A4.6
A/Map position: 4
A/Introns: 9/3; 43/3; 70/3

Query Match 13.4%; Score 182.5; DB 2; Length 317;
Best Local Similarity 43.6%; Pred. No. 8.4e-06;
Matches 51; Conservative 3; Mismatches 44; Indels 19; Gaps 7;
QY 15 GNPNGQGMFGAMGNO-PAGAGGYPGA-SYGYPGQ-APPGAYPQAPPGAYHGA 67
DB 205 GPPGPPGSGA-----PSPGAPGDDAPSEPIITPSPGPPG-PQAGGPDGPGAPG 258
QY 68 AYPGAPAGVYPGPPSGPGAYPSSGQSPAGATGPGYGAAPGLIYPNLP 124
DB 259 G-PGTPGS---KGPNGNPGAPGADGNPGAGQAGTSGAGEKG--ICPKYCAIDGV 309

RESULT 31
T26185
hypothetical protein W05B2.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T26185
R/Gardner, A.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z20166
A/Accession: T26185
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-304 <MIL>
A/Cross-references: EMBL:Z81138; PIDN:CAB03475.1; GSPDB:GN00021; CESP:W05B2.1
A/Experimental source: clone W05B2
C/Genetics:
A/Gene: CESP:W05B2.1
A/Map position: 3
A/Introns: 27/3

Query Match 13.4%; Score 181.5; DB 2; Length 304;
Best Local Similarity 44.8%; Pred. No. 9.4e-06;
Matches 56; Conservative 6; Mismatches 42; Indels 21; Gaps 9;

QY 12 SGSGNP---NPQGMFGAMGNO-PAGAGGYPGA-SYGYPGQAPPGAYPQAPPGAY 62
DB 180 SGAGAPGAPGKAGAPGAGNPGAGPQPSDAGSESSPGAPGQGP-----QSPGP- 233
QY 63 HGAPATGATAPGYPGP--PSG-PGAYPSSGQSPAGATGPGYGAAPGLIYPNLP 119
DB 224 AGSPGA-PGGPGQAGAPGPKSPGAPGQPGADGNPGAGC--QPGSGGAGEKGICPKYCA 290
QY 120 LPGV 124
DB 291 IDGV 295

RESULT 32
T26184
hypothetical protein W05B2.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T26184
R/Gardner, A.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z20166
A/Accession: T26184
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-304 <MIL>
A/Cross-references: EMBL:Z81138; PIDN:CAB03474.1; GSPDB:GN00021; CESP:W05B2.6
A/Experimental source: clone W05B2
C/Genetics:
A/Gene: CESP:W05B2.6
A/Map position: 3
A/Introns: 27/3

Query Match 13.4%; Score 181.5; DB 2; Length 304;
Best Local Similarity 44.8%; Pred. No. 9.4e-06;
Matches 56; Conservative 6; Mismatches 42; Indels 21; Gaps 9;
QY 12 SGSGNP---NPQGMFGAMGNO-PAGAGGYPGA-SYGYPGQAPPGAYPQAPPGAY 62
DB 180 SGAGAPGAPGKAGAPGAGNPGAGPQPSDAGSESSPGAPGQGP-----QSPGP- 233
QY 63 HGAPATGATAPGYPGP--PSG-PGAYPSSGQSPAGATGPGYGAAPGLIYPNLP 119
DB 224 AGSPGA-PGGPGQAGAPGPKSPGAPGQPGADGNPGAGC--QPGSGGAGEKGICPKYCA 290
QY 120 LPGV 124
DB 291 IDGV 295

RESULT 33
T22827
hypothetical protein F57B1.4 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T22827
R/Sims, M.
submitted to the EMBL Data Library, August 1996
A/Reference number: Z19622
A/Accession: T22827
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-356 <MIL>
A/Cross-references: EMBL:Z78064; PIDN:CAB01508.1; GSPDB:GN00023; CESP:F57B1.4
A/Experimental source: clone F57B1
C/Genetics:
A/Gene: CESP:F57B1.4
A/Map position: 5
A/Introns: 65/1; 111/3

Query Match 13.3%; Score 180.5; DB 2; Length 356;
Best Local Similarity 43.0%; Pred. No. 1.3e-05;

collagen col-13 precursor - *Caenorhabditis elegans*
Species: *Caenorhabditis elegans*

QY	13	GSNG---ENQGMFGAMGNC-PAGAGCYPGASTPYGAPPGAYPGAPFG--AYHGAP	66
Db	134	GNGPGSGPGDGGPGAPGAKNGPFGPNNGA--FGAPG-----FGDDASEPTTPGAP	185
QY	67	GAYGAPAPGVYGPSPSGPAPATSSGQBPAPGATVATGPADAGLI-----VEYNLP	121
Db	186	GP-QGTTGP---QGPPGQPGPQGHDPGAPG-PKSGPNNGCGDAGDANPGAGGSGT	239
QY	122	GGV	124
Db	240	GGV	242

RESULT 49

hypothetical protein F55C10.2 - *Caenorhabditis elegans*
T22705
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
C/Accession: T22705
R/Posson, R. submitted to the EMBL Data Library, June 1996
A/Reference number: Z19603
A/Accession: T22705
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-299 <MIL>
A/Cross-references: EMBL:274036; PIDN:CA98486.1; GSPDE:GN00023; CBSP:F55C10.2
C/Genetics:
A:Gene: CBSP:F55C10.2
A:Map position: 5
A:introns: 5/3

Query Match	12.8%	Score 173.5;	DB 2;	Length 299;
Best Local Similarity	43.1%	Pred. No. 3.3e-05;		
Matches 53;	Conservative 3;	Mismatches 42;	Indels 25;	Gaps 9

QY 13 GSGN---PSPQGMFGAMNGO--PAGAGGYTGAASYTPTGGAPPGAYVGGQAPPG--AYHGAP 66
 Db 167 GNGNGTSGSGGQDQPGAPGANKGSPSGENGNGA--PGAPGQ-----PGDAPSEPTTPGAP 218
 QY 67 GAYPAPAPGYTPGPPSGPGAYPSSGQSPGAPGAYATGPGADAPGLI-----VPYNTLP 121
 Db 219 GP-QSTPSP---QGPPGQPGQPGHGDGPAPG--PKGPGNGNGPGQPGADGNPGAPGSGTF 272
 QY 122 GGV 124
 Db 273 GGV 275

RESULT 5C

S70644
 annexin VII - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 13-Aug-1999
 C:Accession: S70644
 R:Srivastava, M.; Zhang-Keck, Z.-Y.; Caohuy, H.; McPhie, P.; Pollard, H.B.
 B:Biochem. J. 316, 729-735, 1996
 A>Title: Novel isoforms of synexin in *Xenopus laevis*: multiple tandem PCQM repeats distributed
 A:Reference number: S70644; MUID:96265035; PMID:8670145
 A:Accession: S70644
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-512 <SR>
 A:Cross-References: EMBL:U01365; NID:G790543; PIDD:AA81845.1; PID:G790544
 C:Superfamily: annexin VII; annexin repeat homology
 F:212-282/Domain: annexin repeat homology <X1>
 F:283-355/Domain: annexin repeat homology <X2>
 F:367-439/Domain: annexin repeat homology <X3>
 F:443-512/Domain: annexin repeat homology <X4>

	Best Local Similarity	31.6%;	Pred. No. 5.9e-05;	
	Matches	73;	Conservative	16; Mismatches 71; Indels 71; Gaps 17
Qy	12	SSGSEPNQ---GMPAMGNOPAGAGGYGA---SYFGYG---	QAPPGA-----Y	53
Db	48	SGGNAAPPGGGYPIYA-GGYPA-PEGYVGGMPSYFAGAPAGGGCGYVAPGAPAYGV		105
Qy	54	PGQAPFGAYHGAFGAYPGAPAPAYVPGSPG---PGAYPSSG---QPS---		95
Db	106	PGYGGPG-FNAPAGGY-GAPNAGGCGVPPAGGYSPGGAPGYGGPGSPSSQSYAGGPGQ		163
Qy	96	-----APGAYATGPGAPAGGLIAPVYPLPGGVPPMLITLITGVENMARI		143
Db	164	MPGQMPGQMPGQALASGYPSGP-APAPG-TPY-----AAMMATQSTIIQAENFD		210
Qy	144	ALDFQRGNDVAFHPPFENENNRV-ITYCNKRLDNNMGREERQSVFPFESRG		193
Db	211	AL--SDAEKLRKAMKFGTDEKPIIDVYANNSND--RQTKQAAPKTAAYG		256

```
Search completed: August 23, 2004, 14:38:57
Job time : 25 secs
```

Query Match 12.8%; Score 173.5; DB 2; Length 512;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:32:27 ; Search time 52 Seconds
(without alignments)
1492.644 Million cell updates/sec

Title: US-09-297-040-4
Perfect score: 1357
Sequence: 1 MAUNFSLHDAISGSGNPNPQ.....ISKLGISGIDILTSASYTMI 246

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

1: SPTRMBL_25:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_ivirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1082	79.7	264	11	08C253
2	695.5	51.3	262	13	090713
3	663.5	48.9	332	13	08QGD9
4	603	44.4	121	4	086175
5	581	42.8	108	4	081XB9
6	580.5	42.8	261	13	07ZSY1
7	444	32.7	139	11	061357
8	314.5	23.2	341	13	09P11
9	299	22.0	322	11	099183
10	284	20.9	320	13	07T119
11	271	20.0	348	13	08QGB1
12	269.5	19.9	310	13	07T118
13	264	19.5	311	4	08WY07
14	260	19.2	349	6	09XSM9
15	258.5	19.0	315	13	07ZTMS
16	258	19.0	343	13	08W98

17	257	18.9	317	6	09XSM8	09xsm8 sus scrofa
18	246.5	18.2	340	13	08W99	08w99 xenopus lae
19	243.5	17.9	326	11	091X74	091x74 mus musculus
20	238.5	17.6	328	13	07ZTB9	07ztb9 xenopus lae
21	237.5	17.5	326	11	08K419	08k419 mus musculus
22	233.5	17.2	308	6	09TUB8	09tub8 oryctolagus
23	230	16.9	308	13	07ZTB8	07ztb8 xenopus lae
24	228.5	16.8	359	4	08TEV1	08tev1 homo sapien
25	223.5	16.5	359	4	09BXC8	09bxc8 homo sapien
26	219.5	16.2	204	13	07SYP0	07syp0 xenopus lae
27	219.5	16.2	319	13	07ZTB6	07ztb6 xenopus lae
28	216.5	16.0	300	11	088353	088353 mus musculus
29	215.5	15.9	180	13	07T120	07t120 brachydanto
30	213	15.7	332	13	08W997	08w997 xenopus lae
31	210.5	15.5	703	12	083467	083467 porcine ade
32	206.5	15.2	162	6	08MNS9	08mns9 ovis aries
33	195.5	14.4	903	16	08ZHF3	08zif3 streptomyce
34	195	14.4	605	13	0801J8	0801j8 salmo salar
35	194	14.3	223	11	08C6H0	08c6h0 mus musculus
36	192.5	14.2	415	11	08CCV9	08ccv9 mus musculus
37	192.5	14.2	463	11	0922A2	0922a2 mus musculus
38	192.5	14.2	463	11	08BP75	08bp75 mus musculus
39	192.5	14.2	485	13	0804G3	0804g3 brachydanto
40	191	14.1	483	13	07J391	07j391 brachydanto
41	190.5	14.0	177	10	039115	039115 aradidopsis
42	189.5	14.0	329	5	044796	044796 caenorhabdi
43	189.5	14.0	526	13	0804G4	0804g4 brachydanto
44	188.5	13.9	285	5	045904	045904 caenorhabdi
45	187.5	13.8	277	5	09NUV1	09nuv1 haemochus
46	187.5	13.8	283	5	09NUV0	09nuv0 haemochus
47	187.5	13.8	313	5	020922	020922 caenorhabdi
48	187	13.8	300	4	096Q87	096q87 homo sapien
49	186.5	13.7	179	10	09M0L8	09m0l8 aradidopsis
50	186.5	13.7	277	10	043678	043678 aradidopsis
51	185.5	13.7	463	11	08VIN2	08vin2 ratcus nov
52	185	13.6	185	10	084TC1	084tc1 oryza sativ
53	185	13.6	1070	2	09APM8	09apm8 myxococcus
54	184	13.6	239	5	018302	018302 caenorhabdi
55	183.5	13.5	744	11	0921S8	0921s8 mus musculu
56	183.5	13.5	744	11	08BGL6	08bgl6 mus musculu
57	183	13.5	239	5	027318	027318 caenorhabdi
58	183	13.5	1884	5	09NHW2	09nhw2 nephila ina
59	182.5	13.4	304	5	020091	020091 caenorhabdi
60	181.5	13.4	304	5	09XVG3	09xvg3 caenorhabdi
61	181.5	13.4	304	5	09U349	09u349 caenorhabdi
62	181.5	13.4	304	5	09U348	09u348 caenorhabdi
63	181.5	13.4	308	5	094620	094620 meloidogyne
64	181.5	13.4	462	5	09NHW3	09nhw3 nephila cla
65	181.5	13.4	907	5	044359	044359 nephila cla

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	264 AA.
08C253			
ID	08C253		
AC	08C253		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Lectin.		
GN	LGALS3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
FM	SEQUENCE FROM N.A.		
RP	STRAIN=NOD;		
RC	MEDLINE=22354683; Pubmed=12466851;		
RX	The PANTOM Consortium.		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		

"Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK089249; BAC0813.1; -.
DR PIR; A45983; A45983.
DR MGI; 67778; Lgal3.
DR GO; GO:0005737; Cytoplasm; IDA.
DR GO; GO:0005515; F:protein binding; IPT.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 264 AA; 27415 MW; 253AF8037143740 CRC64;

Query Match 79.7%; Score 1082; DB 11; Length 264;
Best Local Similarity 78.6%; Pred. No. 1.2e-85;
Matches 209; Conservative 15; Mismatches 20; Indels 22; Gaps 8;

QY 1 MADNPSLHDLASGSGNPNPGMGPAGAGGYPGAGYPG-YPGQAPPGAYPGQAP 59
DB 1 MADNPSLHDLASGSGNPNPGMGPAGAGGYPGAGYPG-YPGQAPPGAYPGQAP 59
QY 60 GAYHG--APGAYPGAPAPGVYPPSPGPGAYPSS-----GQSPAPGAY-----AT 102
DB 60 GAYPGQAPPGAYPGAPAPGVYPPSPGPGAYPSS-----GQSPAPGAY-----AT 102
QY 103 GPYAGPAGPLVYVPLPLPGGVVPRMLITLIGTVKPNRRIALDFQRGNDAVAFH-PRFN 161
DB 119 GPYGPAGPLVYVPLPLPGGVVPRMLITLIGTVKPNRRIALDFQRGNDAVAFH-PRFN 178
QY 162 ENNRVIVCNTRKLDNNMGREROSVFPFESGKPFKIQLVPEPDHFKVAVNDAH-LQYNHR 220
DB 179 ENNRVIVCNTRKLDNNMGREROSVFPFESGKPFKIQLVPEPDHFKVAVNDAH-LQYNHR 238
QY 221 VKKLEISKLGISGDIIDLTSAASYTMI 246
DB 239 MKNLREISQLGSGDITLTSANHAMT 264

RESULT 2

QY 090713 PRELIMINARY; PRT; 262 AA.
DB 090713
AC 090713
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Galectin-3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RA Nurnskaya M.V., Linsenmayer T.F.;
RT "Identification and characterization of up-regulated genes during
RT chondrocyte hypertrophy."
RL Dev. Dyn. 0:0-0 (1996).
DR EMBL; U50339; AAB02856.1; -.
DR HSRF; P17931; IAK3.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 262 AA; 28155 MW; 2D89581493B0CA30 CRC64;

Query Match 51.3%; Score 695.5; DB 13; Length 262;
Best Local Similarity 58.8%; Pred. No. 3.3e-52;
Matches 150; Conservative 23; Mismatches 51; Indels 31; Gaps 14;

QY 7 LHDALSGS-----GNPNPGMGPAGAGGYPGAGYPGYPGQAP--PGAYPGQAP 58
DB 24 LHPQLSDLPAHNPAGAPPQGM-----NRPFGAPF--ATPGYFGAYPGAPGPPG--A 74
QY 59 PGAYHGAPGAYPGAPAPGVYPPSPGPGAYPSSGQSPAPGAY--ATPGY--APAGPLIV 114
DB 75 PGPHHGPPGYPGPGP-PGPYPGPG--PGYPF--GGP--PGYPGPGPTAPYSBAPAPLKV 127
QY 115 PYNLPLPGGVVPRMLITLIGTVKPNRRIALDFQRGNDAVAFH-PRFNENNRVIVCNTRK 173
DB 128 PYDLPPLPAGLMPRLITLITGTVNSPNRNFSLDFRKGQDIAHFPRDFEDHKKRVIVCNRM 187
QY 174 LDNNMGREROSV-FPFESGKPFKIQLVPEPDHFKVAVNDAH-LQYNHRVKKLEISKLG 231
DB 188 FQNMWKEERTAPRPFPFPGPTGPFKLIQLCEGDHFKVAVNDAHLLQFNREKKNIGITLC 247
QY 232 ISGDIIDLTSAASYTMI 246
DB 248 IAGDITLTSVLTSMI 262

RESULT 3

QY 08QGD9 PRELIMINARY; PRT; 332 AA.
DB 08QGD9
AC 08QGD9
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Galectin-3TMI isoform containing transmembrane spanning domain.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RA Gorski J.P., Liu F.-T., Artigues A., Castagna L.F., Osoby P.;
RT "New alternatively spliced form of galectin-3, a member of beta-
RT galactoside-binding animal lectin family, contains predicted
RT transmembrane spanning domain and leucine zipper motif."
RL Submitted (Feb-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF479564; AAL91920.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
KW Transmembrane.
SQ SEQUENCE 332 AA; 35484 MW; CF3203BC99418CB1 CRC64;

Query Match 48.9%; Score 663.5; DB 13; Length 332;
Best Local Similarity 46.5%; Pred. No. 2.6e-49;
Matches 151; Conservative 24; Mismatches 49; Indels 101; Gaps 15;

QY 7 LHDALSGS-----GNPNPGMGPAGAGGYPGAGYPGYPGQAP--PGAYPGQAP 58
DB 24 LHPQLSDLPAHNPAGAPPQGM-----NRPFGAPF--ATPGYFGAYPGAPGPPG--A 74
QY 59 PGAYHGAPGAYPGAPAPGVYPPSPGPGAYPSSGQSPAPGAY--ATPGY--APAGPLIV 113
DB 75 PGPHHGPPGYPGPGP-PGPYPGPG--PGYPF--GGP--PGYPGPGPTAPYSBAPAPLVT 127
QY 114 -----VYNUPLPGGVVPRMLITLIGTVKPNRRIALDFQRGNDAVAFH-PRFNEN 163
DB 188 TPNPCLDKVYVDLPLPAGLMPRLITLITGTVNSPNRNFSLDFRKGQDIAHFENFRFKED 247
QY 164 NNRVIVCNTRKLDNNMGREROSV-FPFESGKPFKIQLVPEPDHFKVAVNDAH-LQYNHRV 221

Db 248 HRRVIVCSNMFQNNWKEERTAPRPFEPETPKLVLCGGHFKYAVNDAHLQNPFE 307
 QY 222 KKLNEISKIGSDIDLTASTMT 246
 Db 308 KKLNEITKLCIAGDITLTSLTSMI 332

RESULT 4

Q86TY5 PRELIMINARY; PRT; 121 AA.
 ID Q86TY5
 AC Q86TY5;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Human full-length cDNA clone CS0D1041YE05 of placenta of Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Genoscope;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Li W.B., Gruber C., Jessee J., Polayes D.;
 RT "Full-length cDNA libraries and normalization";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BX161455; CAD61918.1; -
 DR GO: GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro: IPR008985; Cona like lec_g1.
 DR InterPro: IPR001079; Galectin.
 DR Pfam: PF00337; Gal-bind lectin; 1.
 DR SMART: SM00276; GLECT; 1.
 DR PROSITE: PS00309; GALACTIN; 1.
 KW Plasmid
 SQ SEQUENCE 121 AA; 13897 MW; 020099E4DA6D32FA CRC64;

Query Match 44.4%; Score 603; DB 4; Length 121;
 Best local Similarity 98.3%; Pred. No. 1.3e-44;
 Matches 119; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 128 MLTITIGYKPNRRIALDFQGNDAFHP-PRFNNRRVIVCNTKLDNNWGREEROVY 186
 Db 1 MLTITIGYKPNRRIALDFQGNDAFHP-PRFNNRRVIVCNTKLDNNWGREEROVY 60
 QY 187 FPESGKPFKIVLVEPDHRYKAVNDAH-LQYHRYKCLNEISKLGISGIDLTASAYTM 245
 Db 61 FPESGKPFKIVLVEPDHRYKAVNDAH-LQYHRYKCLNEISKLGISGIDLTASAYTM 120
 QY 246 I 246
 Db 121 I 121

RESULT 5

Q81XB9 PRELIMINARY; PRT; 108 AA.
 ID Q81XB9
 AC Q81XB9;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Galectin 3 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Neugebauer S., Daimon M., Baba T., Watanabe T.;

RT "A common mutation of an AGE-binding protein is not associated with
 RT diabetic microvascular complications in type 2 diabetes."
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB086821; BAC45249.1; -
 FT NON-TER 1
 FT NON-TER 1
 SQ SEQUENCE 108 AA; 10035 MW; 5D2BE70C2CF626C5 CRC64;

Query Match 42.8%; Score 581; DB 4; Length 108;
 Best local Similarity 96.3%; Pred. No. 9.1e-43;
 Matches 104; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 7 LHDALSSGNNPQGWPGWGNOPAGAGYGCASYPG-YPCQAPPGATPGQAPPGAYHGA 65
 Db 1 LHDALSSGNNPQGWPGWGNOPAGAGYGCASYPG-YPCQAPPGATPGQAPPGAYHGA 60
 QY 66 PGAYPGAPPGVYPGPSGPGAYSSGQSPASGAY-ATGPAGAPAGPL 112
 Db 61 PGAYPGAPPGVYPGPSGPGAYSSGQSPASGAY-ATGPAGAPAGPL 108

RESULT 6

Q7ZSY1 PRELIMINARY; PRT; 261 AA.
 ID Q7ZSY1
 AC Q7ZSY1;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Similar to lectin, galactose binding, soluble 3 (Galectin family
 DE xgallectin-VIIA).
 GN XGALLECTIN-VIIA.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Strausberg R.;
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22552404; PubMed=12538594;
 RX Shoji H., Nishi N., Hirashima M., Nakamura T.;
 RA "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY
 RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING
 RT EMBRYOGENESIS.";
 RL J. Biol. Chem. 278:12285-12293 (2003).
 DR EMBL: BC046662; AAH46662.1; -
 DR EMBL: AB080020; BAC55886.1; -
 DR GO: GO:0005529; F:sugar binding; IEA.
 DR GO: GO:0007157; P:heterophilic cell adhesion; IEA.
 DR InterPro: IPR008985; Cona like lec_g1.
 DR InterPro: IPR001079; Galectin.
 DR Pfam: PF00337; Gal-bind lectin; 1.
 DR SMART: SM00276; GLECT; 1.
 DR PROSITE: PS00309; GALACTIN; 1.
 KW Lectin.
 SQ SEQUENCE 261 AA; 28332 MW; EAF78A115BE4DB91 CRC64;

Query Match 42.8%; Score 580.5; DB 13; Length 261;
 Best local Similarity 49.1%; Pred. No. 3e-42;
 Matches 135; Conservative 31; Mismatches 66; Indels 43; Gaps 15;

QY 1 MADNPSLHDALS--GSGNP-----NPQGWPGWGN--QPAGAGYPGA---SYPGYPGQ 47
 Db 1 MADNPSLHDALNQTANPQGSQDPGQPGQGWAMGNRAPGYPGAPPGQYRYPGA 60
 QY 48 AP-----PGAYPGAPPGAYHGA PG-AYGADAPGV-YFG-PPSGGAYPSSGQPSA 96
 Db 61 APGGQGYRFGPGYPGQ-----QYRFGPGQGFPGFAPPGQGYGIGIADQPPQKXGVPGQDEP 116

QY 97 PGAYATGPGAPAGP-LIVPNLPLPGVVPRLITITIGTKPNANRIALDFQGNDA 154
 DB 117 P-----KSPAPVQLKVPYDLPSPGVPRMLITIGTVNPNPREFVADPKRGODIA 167
 QY 155 FHF-PRFNENNRKIVYNTLTDNNNGEEROS-VPEPSGPKPKIQLVPEPDKVAVND 212
 DB 168 LHNPRFDE-RPNIVNNSMIGNMGEERKAPKPFVACQPKIQLVCEADHYKVGINN 226
 QY 213 AHL-QVNRVYKLNLSKLGISGIDILTSASYMTI 246
 DB 227 ENLFQYVRYRRLHLSISVCIIGVTLTDANVTAV 261
 RESULT 7
 Q61357 PRELIMINARY; PRT; 139 AA.
 AC Q61357; 1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBlrel. 25, Last annotation update)
 DE Beta-galactoside-binding lectin (L-34) (Fragment).
 GN LGLS3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88080093; PubMed=3335026;
 RA Raz A., Carmi P., Pazerant G.;
 RT "Expression of two different endogenous galactoside-binding lectins
 sharing sequence homology.";
 RL Cancer Res. 48:645-649 (1988).
 DR EMBL; M33215; AAA37314.1; -;
 DR PIR; S08576; S08576.
 DR HSSP; P17931; IA3K.
 DR MGD; MGI:96778; LgalS3.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005515; F:protein binding; IPT.
 DR InterPro; IPR008985; Cona_like_lect_g1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 1.
 DR SMART; SM00276; GLECT; 1.
 DR PROSITE; PS00309; GALAPTIN; 1.
 KW Lectin.
 KM
 SQ SEQUENCE 139 AA; 15914 MW; B6373FDG378568 CRC64;
 FT NON TER
 Query Match 32.7%; Score 444; DB 11; Length 139;
 Best Local Similarity 76.8%; Pred. No. 9e-31; Indels 2; Gaps 2;
 Matches 86; Conservative 7; Mismatches 17;
 QY 104 PGAPAGPLIVPNLPLPGVVPRLITITIGTKPNANRIALDFQGNDAFHF-PRFNE 162
 DB 22 PIMCPMTTVDVYDLPFGVWPRMLITIGTVKPNANRIALDFQGNDAFHFPRFNE 81
 QY 163 NNRRVIVCNTKLDNNNGEEROSVPPESGPKPKIQLVPEPDKVAVND 214
 DB 82 NNRRILVCMTKDNNNGEEROSVPPESGPKPKIQLVLAADHSGCG-HDAH 132
 RESULT 8
 Q9PT11 PRELIMINARY; PRT; 341 AA.
 AC Q9PT11; 2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Galectin like protein.
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

OX NCBI_TaxID=8022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head kidney;
 RA Soma G., Inagawa H., Honda T., Nakanishi T., Otake M., Nishizawa T.,
 RA Kanou J., Endo M.;
 RT "Rainbow trout cDNA similar to galectin 9 like.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB027452; BAA8670.1; -;
 DR HSSP; P47929; 1BKZ.
 DR InterPro; IPR008985; Cona_like_lect_g1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 2.
 DR SMART; SM00276; GLECT; 2.
 DR PROSITE; PS00309; GALAPTIN; 1.
 SQ SEQUENCE 341 AA; 37816 MW; 176D835BC1950C4 CRC64;
 Query Match 23.2%; Score 314.5; DB 13; Length 341;
 Best Local Similarity 36.0%; Pred. No. 4.4e-19;
 Matches 90; Conservative 30; Mismatches 69; Indels 61; Gaps 17;
 QY 9 DAISSG-----NPN-----QWPGAMGQAGAGVPGASVPGYGGAPGAY 53
 DB 132 DTISADKVELTSTVFONPAPTTPAQGFPA-----QF-----GFP--SYGFPAG--PG-F 178
 QY 54 PGQAPGAYHGAAPGAPGAPGVPYPPGPGGAYPSSGQPSAPGAYATGPY-GAPAGPL 112
 DB 179 P-----SYGFPA--QPGFPSCPG-FP--GGQGF-----PYGFPAGP- 211
 QY 113 TVPNLPLPGVVPRLITITIGTKPNANRIALDFQGNDAFHF-PRFNENNRKIVCN 171
 DB 212 AVPKNMINGLYRGRTINIGVNPANRPHINLENSGIALHNPFFDT--LVARN 268
 QY 172 TKLDNNNGEEROSVPPESGPKPKIQLVPEPDKVAVNDHQLQ-YNHRVYKLNLSK 230
 DB 269 SKLPDQWKEERSSGMPFRHQAFITSTCAQCKIYVNNQNSITYHRTLLQGVNVL 328
 QY 231 GISGIDILTS 240
 DB 329 EVDGLSLTS 338
 RESULT 9
 Q99L83 PRELIMINARY; PRT; 322 AA.
 AC Q99L83; 2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Lectin, galactose binding, soluble 9.
 GN LGLS9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strusberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003754; AAH03754.1; -;
 DR HSSP; P17931; IA3K.
 DR MGD; MGI:109496; LgalS9.
 DR InterPro; IPR008985; Cona_like_lect_g1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 2.
 DR SMART; SM00276; GLECT; 2.
 DR PROSITE; PS00309; GALAPTIN; 2.
 SQ SEQUENCE 322 AA; 36545 MW; 032D77400737562E CRC64;
 Query Match 22.0%; Score 299; DB 11; Length 322;
 Best Local Similarity 37.5%; Pred. No. 8.9e-18;
 Matches 69; Conservative 28; Mismatches 47; Indels 40; Gaps 8;

Db 292 NDIDILEICGDLQLTLS 307

RESULT 13

Q8WYQ7 PRELIMINARY; PRT; 311 AA.
AC Q8WYQ7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Galectin-9.
GN LGALS9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Aktyama S.;
RT "Homo sapiens galectin-9 (LGALS9) / ecalectin gene, exon 2 through
RT 11.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB040130; BAB83623.1; -.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 311 AA; 34690 MW; 3D00232FE39D776C CRC64;

Query Match 19.5%; Score 264; DB 4; Length 311;
Best Local Similarity 36.5%; Pred. No. 9, 1e-15;
Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;
QY 65 ARGATPGAPA- PGVPPGSPGAYPSSGQSPAGAVATGYGAPAGPLVPPNLPFG 122
DB 159 AGQWFTSTAIPTPMYPHP-----AVP-----MPFITTLIG 189
QY 123 GVPPNMLITLITGVNPNRRIALDFQGNDAVHF- PRENNRRVYVNTKLDNNWGRE 181
DB 190 GLYPSKSTLSTVLPASQRPHINLCSGNHIAFHNPRDEN--AVVNTQIDNSGSE 246
QY 182 ERQ--SVFPESGKPEKIVLVEPDHFKVAVNDAH-QYNHRVKLINEISKIGISDIDL 238
DB 247 EESLPRKMFVGGQSPFWILCEAHGLKVAVDGQHLFEYVYHRLNPLTNLEVGGLDIL 306
QY 239 T 239
DB 307 T 307

RESULT 14

Q9XSM9 PRELIMINARY; PRT; 349 AA.
AC Q9XSM9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Urate transporter/channel protein, isoform (UATP.1).
GN UATP.1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Kidney;
RC Soltzenberger F.; Graessler J.; Schroeder H.E.;
RT "Molecular characterization and functional expression of a renal urate
RT transporter/channel from cultured LLC-PK1 epithelial cells.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131827; CAB44279.1; -.

DR HSSP; P17931; 1A3X.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 349 AA; 38899 MW; BF83D3E213E7B64C CRC64;

Query Match 19.2%; Score 260; DB 6; Length 349;
Best Local Similarity 31.7%; Pred. No. 2, 3e-14;
Matches 66; Conservative 32; Mismatches 68; Indels 42; Gaps 9;
QY 37 PGASYPGPGQ--APGAYPGQAPP-----GAYHGAAGAYGAPAGVYPPSGPGAY 88
DB 167 PACPPPHKGRKRPGRWPNNSAPITQVTHVQSTPGQ-----MFNPMPIMPMAY 218
QY 89 PSSGQSPAGAVATGYGAPAGPLIVPNLELPGGVPPNMLITLITGVKPNANRIALDFQ 148
DB 219 FN-----PVFP--IPFPASIPGLYPSKISWSTLTPSOSTFYINLR 259
QY 149 RGNDAVHF- PRENNRRVYVNTKLDNNWGREER--QSVFPESGKPEKIVLVEPDH 205
DB 260 GSDIAFHLNPRENEN--AVVNTQIDSSWGPBERGLPRKMPFPGQSLWLILCESHC 316
QY 206 FKVAVNDAH-QYNHRVKLINEISKIGI 232
DB 317 FKVAVDQGHLEFYVYHRLKHLPTINSLEV 344

RESULT 15

Q7ZTB5 PRELIMINARY; PRT; 315 AA.
AC Q7ZTB5;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Galectin Family xgalectin-VIIIA.
GN XGALECTIN-VIIIA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-22552404; PubMed-12538594;
RA Shoji H.; Nishi N.; Hiraishina M.; Nakamura T.;
RT "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY
RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING
RT EMBRYOGENESIS.";
RT RT
RT J. Biol. Chem. 278:12285-12293 (2003).
DR EMBL; AB080021; BAC55887.1; -.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
SQ SEQUENCE 315 AA; 35895 MW; 3A42275840579A9A CRC64;

Query Match 19.0%; Score 258.5; DB 13; Length 315;
Best Local Similarity 43.2%; Pred. No. 2, 8e-14;
Matches 60; Conservative 21; Mismatches 47; Indels 11; Gaps 5;
QY 113 IVPYNIPLPGVPPNMLITLITGVKNANRIALDFQGN-----DVAFHF- PRENNNR 165
DB 14 VEVYVGTITIGLELPGQIVIHGVHDADRFQDFQGNQSVQPSDVAFHFNFQSGSH 73
QY 166 RVATVNTKLDNNWGREERQSVFPESGKPEKIVLVEPDHFKVAVNDAH-QYNHRVKYL 224
DB 74 --IVCNTLENEKMGWEKTYQMPFTYQGPFEIILFVFDHDFQVSSGNKMLLVYHRI-SL 130
QY 225 NEISKIGISDIDLITASY 243
DB 131 QRVDTLIGISGVKINTIGF 149

RESULT 16

Q8UM98

ID Q8UM98 PRELIMINARY; PRT; 343 AA.

AC Q8UM98; 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Galectin family xgalectin-IIa.

GN XGALECTIN-IIA.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;

OC Xenopodinae; Xenopus.

NCBI_TaxID=8355;

RN NCBI_TaxID=8355;

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Shoji H., Nishi N., Hirashima M., Nakamura T.;

RT "Purification and cDNA cloning of Xenopus liver galectins, and their

expression in adult tissues."

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB060971; BAB83258.1; -

DR InterPro; IPR008985; Cona like_1ec_g1.

DR InterPro; IPR001079; Galectin.

DR Pfam; PF00337; Gal-bind_lectin; 2.

DR SMART; SM00276; GLECT; 2.

SQ SEQUENCE 343 AA; 38361 MW; B55056ED8F063589 CRC64;

Query Match 19.0%; Score 258; DB 13; Length 343;
Best Local Similarity 34.7%; Pred. No. 3.4e-14;

Matches 74; Conservative 23; Mismatches 86; Indels 30; Gaps 8;

QY 34 GGYPGASYVGYPGQAP--PGAYPGQAPPGAYHG--APAGYGPAPPGYVPGPSGPGAYP 89

DB 148 GGFPSQCPGAPMPSPPQPGVAPQAFPPGVAFNPPQPGVAPQAFPPGVAFP-----Q 200

QY 90 SSGGSAAGAYATGYGAPAGPLVYVNLPLPGGVPEMLITLIGTVKPNANRIALDFOR 149

DB 201 PSSAPFQQAEE-----IPYQTNLYGGIFPSKTIIVITGVYANPKRHINKF 248

QY 150 GNDVAFHF--PRENNRRVIVCNKLDNNMGERERQ--SVFPSSGKPFKIQVLEPDHF 206

DB 249 HGGTALHFNRFDE---CAIVRSHLNGSWGKEERDLPSCGCFVPGQSFWLQIRCEQDAF 305

QY 207 KVAVNDALH--QYNRVKKLNISKLGISGIDL 238

DB 306 KVMNNGTQICEFHREHNLQIDITLIVGDVYL 338

RESULT 17

Q9XSM8

ID Q9XSM8 PRELIMINARY; PRT; 317 AA.

AC Q9XSM8; 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Urate transporter/channel protein (UATP).

GN UATP.

OS Sus scrofa (pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

NCBI_TaxID=9823;

RN NCBI_TaxID=9823;

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RA Splizenberger F., Graessler J., Schroeder H.E.;

RT "Molecular characterization and functional expression of a renal urate

transporter/channel from cultured LLC-PK1 epithelial cells."

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; A0131826; CAB44278.1; -

DR HSSP; P17931; IA3K.

DR InterPro; IPR008985; Cona like_1ec_g1.

DR InterPro; IPR001079; Galectin.

DR Pfam; PF00337; Gal-bind_lectin; 2.

DR SMART; SM00276; GLECT; 2.

DR PROSITE; PS00309; GALAPTIN; 1.

SQ SEQUENCE 317 AA; 35342 MW; AA09562AC87B8B80 CRC64;

Query Match 18.9%; Score 257; DB 6; Length 317;
Best Local Similarity 32.5%; Pred. No. 3.8e-14;

Matches 63; Conservative 30; Mismatches 61; Indels 40; Gaps 8;

QY 49 PGAYPGQAP-----GAYHGAPGAPGAPGAPGYPGPPSPGSGQSPAPGAYAT 102

DB 149 PPGKRPANSAPEITQIVHVGSTPQ-----MFPNMPKAYPN----- 188

QY 103 GPYGAPAGPLVYVNLPLPGGVPEMLITLIGTVKPNANRIALDFORGNDAFHF--PREN 161

DB 189 -----VPEP--IPFPASIPGILYPSKSIWVSSTILPSAQSFYINLRSGSDIAFHLNPRK 241

QY 162 ENNRKRVIVCNKLDNNMGERER--QSVFPSSGKPFKIQVLEPDHFKYAVNDALH--QYN 218

DB 242 EN--AVVRNTQIGSSWGPEERGLPRKMPFSRGSFLTWIICESHCFKVAVDGHLFEY 298

QY 219 HRVKKLNISKLG 232

DB 299 HRLKHLPTNLSLV 312

RESULT 18

Q8UM99

ID Q8UM99 PRELIMINARY; PRT; 340 AA.

AC Q8UM99;

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Galectin family xgalectin-IIa.

GN XGALECTIN-IIA.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;

OC Xenopodinae; Xenopus.

NCBI_TaxID=8355;

RN NCBI_TaxID=8355;

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Shoji H., Nishi N., Hirashima M., Nakamura T.;

RT "Purification and cDNA cloning of Xenopus liver galectins, and their

expression in adult tissues."

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB060970; BAB83257.1; -

DR InterPro; IPR008985; Cona like_1ec_g1.

DR InterPro; IPR001079; Galectin.

DR Pfam; PF00337; Gal-bind_lectin; 2.

DR SMART; SM00276; GLECT; 2.

SQ SEQUENCE 340 AA; 37228 MW; 6E3A88DEF1177A9D CRC64;

Query Match 18.2%; Score 246.5; DB 13; Length 340;
Best Local Similarity 29.3%; Pred. No. 3.3e-13;

Matches 76; Conservative 32; Mismatches 98; Indels 53; Gaps 11;

QY 24 GAWGNQPAAGAGGP-----GASYPRYPGAP-----PG 51

DB 82 GSWGSEKKKDSFFPKKGSFELAFINSSFEITVNGSSFFYFKHMPLEKRVNSIQVSG 141

QY 52 AYPGQAPGAPGAPGAPGAPG--VYPGPPSGGAYPSSGQSPAPGAYATGPGAP 108

DB 142 DVSVCALITAGGGGGGMAAAPAPAPSPHPPGGMAP--MFPNGGVMPPLPYTAGNLPVM 200

QY 109 AGPLIVPNLPLP-----GVVPEMLITLIGTVKPNANRIALDFORG--NDVAFHF--PR 159

DB 201 GGFV--YNPPIPYGTIQGSGMSPKRTVIRIGTIEGAFRPHINFCAGSSNDVALHFNR 257

QY 160 FNNNRKRVIVCNKLDNNMGERERQ--SVFPSSGKPFKIQVLEPDHFKYAVNDALH--QY 217

```

Db      258 LITENS---VYVNSRLGSGWGGERQLSYNPPRPGQYFDISLRGMDRFTYFVNGQRCDF 314
QY      218 NHRVYKLINEISKLIGSGDI 236
Db      315 AHRYSMLQMDIRIBENAV 333

RESULT 19
Q91X74  PRELIMINARY; PRT; 326 AA.
AC      Q91X74
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Lectin, galactose binding, soluble 4 (Galectin-4) (Hypothetical
        protein).
GN      LGALS4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Colon;
RA      Strausberg R.;
RN      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RP      [2]
RC      SEQUENCE FROM N.A.
RA      STRAIN=BALB/c; TISSUE=Liver;
RA      Maly P., Jenikova G., Cummings R.D.;
RT      "Molecular cloning and tissue distribution of mouse galectin-4.";
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC011236; AAH11236.1; -
DR      EMBL; BC021632; AAH21632.1; -
DR      EMBL; BC030297; AAH30297.1; -
DR      EMBL; AY044870; AAK97790.1; -
DR      MGD; MGI:107536; LgalS4.
DR      InterPro; IPR008985; Cona_like_1ec_g1.
DR      InterPro; IPR001079; Galectin.
DR      Pfam; PF00337; Gal-bind_lectin; 2.
DR      PROSITE; PS00309; GALAPTIN; 2.
KW      Hypothetical protein
SQ      SEQUENCE 326 AA; 36372 MW; 7F3DD89862A851B5 CRC64;

Query Match      17.9%; Score 243.5; DB 11; Length 326;
Best Local Similarity 35.8%; Pred. No. 5.8e-13;
Matches 54; Conservative 31; Mismatches 57; Indels 9; Gaps 5;

QY      99 AYATGYGAPAGPLIVPYNLPLPGGVVPRMLITLIGTVKPNARIALDF-----QRGNDVA 154
Db      2 AYVPAAGYQFTYNPPLPYKRPFGSLVSGSVYIQGAKENMRFFHVNPAVSGDDADVA 61
QY      155 FHF-PPFENNRVIVCNTKLDNNWGSEERQSVFPESGKPFKIQVLVEPDHFKVAVN-D 212
Db      62 FHNPNPFDGDK--VFVNINQSGQMGKEKKKMPFGKGFELVFMVMEHKKVYVNGN 119

QY      213 AHLQYNHVKLINEISKLIGSGDIDLTASY 243
Db      120 SFYEYGHRL-PYQWVTHLQVDGLLEQSLNF 149

RESULT 20
Q7ZTB9  PRELIMINARY; PRT; 328 AA.
AC      Q7ZTB9
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Galectin family xgalectin-11b.
GN      XGALACTIN-11b.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC      Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC      Xenopodinae; Xenopus.
OX      NCBI_TaxID=8355;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22552404; PubMed=12538594;
RA      Shoji H., Nishi N., Hirashima M., Nakamura T.;
RT      "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY
        DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING
        EMBRYOGENESIS.";
RT      EMBOGENESIS";
RL      J. Biol. Chem. 278:12285-12293 (2003).
DR      EMBL; AB080016; BAC55882.1; -
DR      InterPro; IPR008985; Cona_like_1ec_g1.
DR      InterPro; IPR001079; Galectin.
DR      Pfam; PF00337; Gal-bind_lectin; 2.
DR      SMART; SM00276; GLECT; 2.
SQ      SEQUENCE 328 AA; 36194 MW; F70DF7556ABEFA0 CRC64;

Query Match      17.6%; Score 238.5; DB 13; Length 328;
Best Local Similarity 28.8%; Pred. No. 1.6e-12;
Matches 75; Conservative 33; Mismatches 91; Indels 61; Gaps 11;

QY      24 GAWGQAPGAGYF-----GASYPGYQGAP-----PG 51
Db      82 GSWGSEERKQDFFPHKKSFLTFMTNSSSFELTNGSSFYKFRHRLPLBRVDSLOING 141
QY      52 AYPGAPPGAYHGAFAVYGPAP-AP-----GYVPGPSGAYVSSQPSAFAVATGP 104
Db      142 DVTYQSLTIAGGGGGGMAAPMAFMFPNGGVNPLPVPFPAAGNLFVWGGP-----V 192
QY      105 YGAPAGPLIVPYNLPLPGGVVPRMLITLIGTVKPNANIALDFQGG--NDVAFHF-PPFN 161
Db      133 YNP-----VPYNGTIOGLPRKTVTRGFLPEGAEFHNFKAGSSNDLALHNPRLT 247
QY      162 ENNRVIVICNTKLDNNWGSEERQ-SVFPESGKPFKIQVLVEPDHFKVAVNDAHL-QYNNH 219
Db      248 EN--AVVNSQLGHWGGERQLSYNPPRAGQYDILRGMDRFTYFVNGQRCDFAH 304
QY      220 RVKLINEISKLIGSGDIDLT 239
Db      305 RYSMCMIDRIEVEGNVVLVS 324

RESULT 21
Q8K419  PRELIMINARY; PRT; 326 AA.
AC      Q8K419
DT      01-OCT-2002 (TrEMBLrel. 22, Created)
DT      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Galectin-4.
GN      LGALS4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6; TISSUE=Colon;
RA      Hokama A., Tanaka Y., Mizoguchi A.;
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF510729; AAW44060.1; -
DR      MGD; MGI:107536; LgalS4.
DR      InterPro; IPR008985; Cona_like_1ec_g1.
DR      InterPro; IPR001079; Galectin.
DR      Pfam; PF00337; Gal-bind_lectin; 2.
DR      SMART; SM00276; GLECT; 2.
DR      PROSITE; PS00309; GALAPTIN; 2.
SQ      SEQUENCE 326 AA; 36386 MW; 7F3C899862AA23B5 CRC64;

Query Match      17.5%; Score 237.5; DB 11; Length 326;
Best Local Similarity 35.1%; Pred. No. 1.9e-12;
Matches 53; Conservative 31; Mismatches 58; Indels 9; Gaps 5;

```

```

QY 99 AYATGYGAPAGPLIVPNILPFGVVPRLITIGTVKPNANRIALDF----QKGNDA 154
DB 2 AAVPAGGYOPTNPLTPYKPIIPGGLSVGWSVYIQMAKENMRPFVNFVAVGDDGADVA 61
QY 155 FHF-PFENNRRVIVCNTKLDNNWREERQSVFPESGPKFKIQLVDPHFKAIVN-D 212
DB 62 FHFNPFDGWDK--VFETKMGQGWKEKKSMPPQKHFELVPMVMEHYKVVNNGN 119
QY 213 AHIQYNHRYKLNLSKLGISGDIIDLTASV 243
DB 120 SFVEYGHRL-PVQMTLHLQVDDDELQSLINF 149

RESULT 22
Q9TUB8 PRELIMINARY; PRT; 328 AA.
AC Q9TUB8
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Galectin-4.
GN LGALS4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New England White;
RX MEDLINE=20147368; PubMed=10683764;
RA Jiang W., Puch S., Guo X., Bhavanandan V.P.;
RT "Signature sequences for the galectin-4 subfamily.";
RL IUBMB Life 48:601-605(1999).
DR HSSB; P47929; IRLKZ.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 328 AA; 36462 MW; 030BF1B723D2ACF CRC64;

Query Match 17.2%; Score 233.5; DB 6; Length 328;
Best Local Similarity 34.5%; Pred. No. 4.3e-12;
Matches 68; Conservative 23; Mismatches 81; Indels 25; Gaps 10;

QY 55 GOAPGAYHAGAPAGAPVYRGPSPGAYSSQPSAPGAYATGPGAPAGPLIV 114
DB 152 GQPTP--HQPMTGPGSPSPY--GHP-GYGAQGLHSLPSMEGPTPNP-----PV 197
QY 115 PYNLEPLPGVVPRLITIGTVKPNANRIALDFQGN--DVAFHF-PFENNRRVIVCN 171
DB 198 PFTGLQGGILARILIVKGVPRIGKFLINFKVSGGDLAHINPMTEG---VVVRN 254
QY 172 TKLDNNWREERQSVF-PESGPKFKIQLVDPHFKAIVNDAH-LQYNHRYKLNLSK 229
DB 255 SRLNGSWGAEERKKAIVNPFQYVDFLSIRCGMDRFKYANGQLFDYAHFFPAQKVDV 314
QY 230 LGISGDIIDLTASVYMI 246
DB 315 IEIQGDVAL--SVQI 328

RESULT 23
Q9TUB8 PRELIMINARY; PRT; 308 AA.
AC Q9TUB8
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB Galectin family xgalectin-11B.
GN XGALECTIN-11B.

```

```

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22552404; PubMed=12538594;
RA Shoji H., Nishi N., Hirashima M., Nakamura T.;
RT "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY
RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING
RT EMBRYOGENESIS.";
RL J. Biol. Chem. 278:12285-12293 (2003).
DR EMBL; AB080017; BAC55883.1; -.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
SQ SEQUENCE 308 AA; 34808 MW; A1C22D04079048AE CRC64;

Query Match 16.9%; Score 230; DB 13; Length 308;
Best Local Similarity 37.6%; Pred. No. 7.9e-12;
Matches 53; Conservative 30; Mismatches 48; Indels 10; Gaps 6;

QY 105 YGAPAGPLIVPNILPFGVVPRLITIGTVKPNANRIALDFQGN--NDVAFHF-PR 159
DB 3 FGAINN-PVPFTLPLSHGDHGLVLSGAVTHSGDFNFNFQCGHSNDIAHFENR 61
QY 160 FENNRRVIVCNTKLDNNWREERQSVFPESGPKFKIQLVDPHFKAIVNDAH-LQYN 218
DB 62 FIDG--IVVCNTERQSGWKEENKREMPFHHQPEFRIILVTHNSYVSVNRNHYLVH 119
QY 219 HRVKLNLSKLGISGDIIDLT 239
DB 120 HRI-PIQRVNTMTLIGCVALT 139

RESULT 24
Q8TEV1 PRELIMINARY; PRT; 359 AA.
AC Q8TEV1
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Lymphocyte/NHL galectin-8 long isoform.
GN LGALS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Moisan S., Mercier J., Demers M., Belanger S.D., Alain T.,
RA Kosakowska A.E., Potworowski E.F., St-Pierre Y.;
RT "Galectins in murine and human non-Hodgkin's lymphomas.";
RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF468213; AAL77076.1; -.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 359 AA; 40151 MW; F87EBFE92E46F571 CRC64;

Query Match 16.8%; Score 228.5; DB 4; Length 359;
Best Local Similarity 40.0%; Pred. No. 1.3e-11;
Matches 56; Conservative 22; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVPYNLPLPGVVPRLITIGTVKPNANRIALDFQGN-----DVAFHF-PFENNRR 165
DB 16 VIVVYVIRPQULDPGTVIVICGVPSDADFQVLDQNGSSVKPRADVAHFHNPFKAG- 74
QY 166 RVIVCNTKLDNNWREERQSVFPESGPKFKIQLVDPHFKAIVNDAH-LQYNHRYK 224

```

```

Db      75 -CIVCCTLLINEKMGREBEITYDTPFKREKSFELIVMWLKDQFVAVNGKATLLYGRIGP- 132
QY      225 NEISKLGISGIDIDLTASYSY 244
      133 EKIDTLGIYGVKNHISIGFS 152
Db
RESULT 25
Q9BXC8 PRELIMINARY; PRT; 359 AA.
AC Q9BXC8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Colorectal carcinoma-derived galectin-8 variant II.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Lahm H., Siebert H.-C., Andre S., Hoeflich A., Diehl D., Sordat B.,
RA Kallner H., Wolf E., Gabius H.-J.;
RT "Coca (colorectal carcinoma-derived) galectin-8 variant II."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF342816; AKL1636.1; -.
DR HSSP; P17931; 1A3K.
DR InterPro; IPR008985; Cona1ike_1ec_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-blind lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 359 AA; 40363 MW; 39BBD61411FAD98 CRC64;

Query Match 16.5%; Score 223.5; DB 4; Length 359;
Best Local Similarity 39.3%; Pred. No. 3.5e-11;
Matches 55; Conservative 23; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVEYNLEPLPGVYPRMLITLLGTGVKNANRIALDPQRG-----DVAFHF-PRFENNRR 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 VIFPVGITIPQDLPGTLIVRGHVPSDADFQVDLNGSSMKPRADVAEFHFNRFYRAG- 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 166 RVIVCTKLDNMNGRRRQGVFPFESGKPKKIQLVVEPDHFKAANDAH-LOYNHRVKKL 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 -CIVCCTLLINEKMGREBEITYDTPFKREKSFELIVMWLKDQFVAVNGKATLLYGRIGP- 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 NEISKLGISGIDIDLTASYSY 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 EKIDTLGIYGVKNHISIGFS 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 26
Q7SYPO PRELIMINARY; PRT; 204 AA.
AC Q7SYPO;
DT 01-OCT-2003 (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S.J., Strusberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus

```

```

RT intrative."
RL Dev. Dyn. 225:384-391 (2002).
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22188257; PubMed=12477932;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Warr M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S., Strusberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054324; AAH54324.1; -.
KW Hypothetical protein.
FT
SQ SEQUENCE 204 AA; 22637 MW; 82532249A1339BE1 CRC64;

Query Match 16.2%; Score 219.5; DB 13; Length 204;
Best Local Similarity 33.3%; Pred. No. 3.9e-11;
Matches 57; Conservative 25; Mismatches 74; Indels 15; Gaps 7;

QY 73 PARGVYPRPSPRGAVPSSGQSPAPGAVATGYGAPAGLIYPNLEPLPGVYPRMLIT 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 39 PASGA--GGAKGAGLIMSTQENLPW----LGPILLPLIPFKAMITGMIIPKTVLM 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 LGTVKPNANRIALDPQRG--NDVAFHF-PRFENNRRVIVCTKLDNMNGRRRQGV-FP 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 KGIIVNNAKNKQISFRVGTINDIALHINPLKNN--TLIRNSFINGWGECKDYVKNP 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 FESGKPEFKIQLVVEPDHFKAANDAH-LOYNHRVKKLNEISKLGISGIDIL 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 FHQGEHFDISIRSGEKQKYVYVNGYHCFNYPHRLTNLQCVDTLEADGDIXL 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 27
Q7ZTB6 PRELIMINARY; PRT; 319 AA.
AC Q7ZTB6;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Galectin family xgalectin-VIIA.
GN XGALCTIN-VIIA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22552404; PubMed=12538594;
RA Shoji H., Nishi N., Hiraehima M., Nakamura T.;
RT "Characterization of the Xenopus Galectin Family THREE STRUCTURALLY
RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING

```

```

RT EMBRYOGENESIS."
RL J. Biol. Chem. 278:12285-12293 (2003).
DR EMBL; AB080019; BAC55885.1; -.
DR InterPro; IPR008985; Cona_1ike_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
SQ SEQUENCE 319 AA; 35369 MW; 2C12918A64B1EBF CRC64;

Query Match
Best Local Similarity 16.2%; Score 219.5; DB 13; Length 319;
Matches 57; Conservative 25; Mismatches 74; Indels 15; Gaps 7;

QY 73 PARGVPPPPGPGAYPPSSGPPAPGAYATGPGYAGAPGLIVPNLPLPGGVPRMLITI 132
DB 154 PASGA--GGAKGAGLMSSTGENTLPPM-----LGPILLHPLIPFKAMIPGMIPIKRYIM 206
QY 133 LGTVPNRNRIALDQFG--NDVAFHF--PRENNRRVIVCTKLDNNWGREEROSV-FP 188
DB 207 KGLVNSNAKNFQISFRKVGVTNDIALHINPRLNK--TLIRNSFNGTWGESEKDVVKNP 263
QY 189 FESGKPFKIQVLVEPDHFKAIVNDAAH-LQYNRVKKLNEISKLGISGIDL 238
DB 264 FHQGEHFDISTRSSEKQKVVYNGYHCNYPHRLNTLQGVDTLEADGDIKL 314

RESULT 28
ID 088353 PRELIMINARY; PRT; 300 AA.
AC 088353;
DT 01-NOV-1998 (TREMBLrel. 08. Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Galectin-4 (Fragment).
GN LGALS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=COLON;
RX MEDLINE=98112847; PubMed=9446608;
RA Gilt M.A., Colnot C., Poltier F., Nani K.J., Barondes S.H.,
RA Leffler H.;
RT "Galectin-4 and galectin-6 are two closely related lectins expressed
RT in mouse gastrointestinal tract."
RL J. Biol. Chem. 273:2961-2970 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=INTESTINE;
RX MEDLINE=98112848; PubMed=9446609;
RA Gilt M.A., Xia Y.-R., Atchison R.B., Luis A.L., Barondes S.H.,
RA Leffler H.;
RT "Sequence, structure, and chromosomal mapping of the mouse Lgal6
RT gene, encoding galectin-6."
RL J. Biol. Chem. 273:2961-2970 (1998).
RN [1]
RP SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: EPITHELIAL CELLS OF THE EMBRYONIC AND ADULT
CC GASTROINTESTINAL TRACT. EXPRESSED AT ABOUT EQUAL LEVELS IN COLON
CC AND SMALL INTESTINE BUT MUCH LESS IN STOMACH.
CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; AF026795; AAC27245.1; -.
DR HSSP; P17931; IAK3.
DR MGD; MGI:107536; Lgal6.
DR GO; GO:0005534; F:galactose binding; IEA.
DR GO; GO:0005529; F:sugar binding; IEA.
DR GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR InterPro; IPR008985; Cona_1ike_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.

```

```

DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galectin; Lectin; Repeat.
FT NON TER 1
FT DOMAIN 1
FT 128
FT 129 176 GALAPTIN 1.
FT 177 >300 LINKER.
FT BINDING 63 69 GALAPTIN 2.
FT BINDING 238 244 BETA-GALACTOSIDE (BY SIMILARITY).
FT NON TER 300 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 300 AA; 33314 MW; 79A80AC8A945AC33 CRC64;

Query Match
Best Local Similarity 16.0%; Score 216.5; DB 11; Length 300;
Matches 47; Conservative 30; Mismatches 45; Indels 9; Gaps 5;

QY 119 PLPGGVPRMLITIGVKNRNRIALDF---QKNDVAFHF--PRENNRRVIVCTNK 173
DB 1 PIPGGLSVGMGVYIQGMAKENRRFHVNFVAGQDDGADVAFHFNRFDGMDK--VFXTM 58
QY 174 LDNNWGREEROSVFPFESGKPFKIQVLVEPDHFKAIVN-DALQYNRVKKLNEISKLG 232
DB 59 QSGQMGKEKKKSNPFQKHFELVFWVPEHYKVVVNGNSFYEYGHRL-DVQVTHLQV 117
QY 233 SGDIDLTSASY 243
DB 118 DGDLELQGINF 128

RESULT 29
ID 077120 PRELIMINARY; PRT; 180 AA.
AC 077120;
DT 01-OCT-2003 (TREMBLrel. 25. Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE SI:d12287.3 (Novel protein similar to vertebrate galectins)
DE (Fragment).
GN SI:D12287.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Babage A.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL954673; CAB17637.1; -.
FT NON TER 180
SQ SEQUENCE 180 AA; 20125 MW; 094C89E259CC8220 CRC64;

Query Match
Best Local Similarity 15.9%; Score 215.5; DB 13; Length 180;
Matches 48; Conservative 33; Mismatches 56; Indels 13; Gaps 6;

QY 100 YATPGYAGAPGLIVPNLPLPGGVPRMLITIGVKNRNRIALDFQGN---DYAF 155
DB 4 YQGPFFYNR---IPSGPIQGLDQGKSIISGRVLPGNRRHVNIQCSHSEANVAL 59
QY 156 HF--PRENNRRVIVCTKLDNNWGREEROSVFPFESGKPFKIQVLVEPDHFKAIVNDAAH 214
DB 60 HFNRVYDPTD--YVNNSYLNRTWGSBERKVESPFAGQGFITGLIIVAGPRYKSTNGRH 117
QY 215 -LQYNRVKKLNEISKLGISGIDLTSASY 243
DB 118 FMDYKRI-PFTQVDTVAVEGMVNSVAF 146

RESULT 30
ID 08UW97 PRELIMINARY; PRT; 332 AA.
AC 08UW97;

```

DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Galectin family xgalectin-IVA.
 GN XGALECTIN-IVA.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodidae; Xenopus.
 NCBI_TaxID=8355;
 RX NCB1
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Shoji H., Nishi N., Hirashina M., Nakamura T.;
 RT "Purification and cDNA cloning of Xenopus liver galectins, and their
 expression in adult tissues."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB060972; BAB83259.1; -
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 2.
 DR SMART; SM00276; GLECT; 2.
 SQ SEQUENCE 332 AA; 36822 MW; D273FD35484B9C2 CRC64;

Query Match 15.7%; Score 213; DB 13; Length 332;
 Best Local Similarity 37.3%; Pred. No. 2.6e-10;
 Matches 50; Conservative 25; Mismatches 53; Indels 6; Gaps 5;

QY 114 VYNNLPFGVVRMLITLITGVKPNANRIALDFQ-RGNDVAFHF-PRENNNRRIYVGN 171
 AC Q83467; PRELIMINARY; PRT; 703 AA.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Porcine adenovirus 4 putative fiber protein.
 OS Porcine adenovirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 NCBI_TaxID=35267;
 RX NCB1
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAV-4;
 RA MEDLINE=96435006; PubMed=8837892;
 RA Kleiboecker S.B.;
 RT "Sequence analysis of the fiber genomic region of a porcine adenovirus
 predicts a novel fiber protein."
 RL Virus Res. 39:299-309(1995).
 DR EMBL; U25120; AAB40001.1; -
 DR HSSP; P17931; 1A3K.
 DR GO; GO:0008637; Picell recognition; IEA.
 DR InterPro; IPR000939; Adeno_fiber2.
 DR InterPro; IPR000978; Adeno_fiber_knob.
 DR InterPro; IPR000931; Adeno_fiber.
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR001079; Galectin.
 DR InterPro; IPR008982; Viral_att.
 DR InterPro; IPR009013; Viral_att_shift.
 DR Pfam; PF00541; adeno_fiber; 1.
 DR Pfam; PF00608; adeno_fiber2; 1.
 DR Pfam; PF00337; Gal-bind_lectin; 2.

RESULT 31

Q83467 PRELIMINARY; PRT; 703 AA.
 AC Q83467;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Porcine adenovirus 4 putative fiber protein.
 OS Porcine adenovirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 NCBI_TaxID=35267;
 RX NCB1
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAV-4;
 RA MEDLINE=96435006; PubMed=8837892;
 RA Kleiboecker S.B.;
 RT "Sequence analysis of the fiber genomic region of a porcine adenovirus
 predicts a novel fiber protein."
 RL Virus Res. 39:299-309(1995).
 DR EMBL; U25120; AAB40001.1; -
 DR HSSP; P17931; 1A3K.
 DR GO; GO:0008637; Picell recognition; IEA.
 DR InterPro; IPR000939; Adeno_fiber2.
 DR InterPro; IPR000978; Adeno_fiber_knob.
 DR InterPro; IPR000931; Adeno_fiber.
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR001079; Galectin.
 DR InterPro; IPR008982; Viral_att.
 DR InterPro; IPR009013; Viral_att_shift.
 DR Pfam; PF00541; adeno_fiber; 1.
 DR Pfam; PF00608; adeno_fiber2; 1.
 DR Pfam; PF00337; Gal-bind_lectin; 2.

DR PRINTS; PF00307; ADENOVSFIBRE.
 DR SMART; SM00276; GLECT; 2.
 SQ SEQUENCE 703 AA; 76649 MW; C2D9E7AEAB3C0E1 CRC64;
 Query Match 15.5%; Score 210.5; DB 12; Length 703;
 Best Local Similarity 32.0%; Pred. No. 1.1e-09;
 Matches 71; Conservative 27; Mismatches 101; Indels 23; Gaps 9;

QY 37 PGASVPGVPGQAPPPAYGQAPPGAYHGAPGA-YRGAPAPGVPPPPSG-----PGAYPS 90
 DB 301 PLASAGVFGKLAASSEMPAPPEAQODQAEEPPAPAEAPAPAEAPAEAPPPRKP 360
 QY 91 SGQSPAPAY-----ATGPGAPAGPLIV-----PYNNLPFGVVRMLITLITGVKPN 140
 DB 361 RGDLAA--LYNRVHSDTRABDPTSPPELVTLDPDFVLPFGVPTGASIVLEGTLTPSA 418
 QY 141 NRIALDFQRG-NDVAFHF-PRENNNRRIYVGNL-DNNNGREEROSVFPESGKPEKI 197
 DB 419 VFTLIDLVTGPAISLALHFNVRPLLEGKHIYVNSREGSSNMGEEVAPPPPPREKPFVL 478
 QY 198 QVLEVPDHFKVAVNDAHLYNNHRVKKINEISKLIGSDIDLTF 239
 DB 479 VIVIGSDTYQITVNGKPLV--DPPQRLQGITRASLSGDLVFT 518

RESULT 32

Q8KN59 PRELIMINARY; PRT; 162 AA.
 ID Q8KN59
 AC Q8KN59;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Galectin-14.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 NCBI_TaxID=9940;
 RX NCB1
 RP SEQUENCE FROM N.A.
 RA Dunphy J.L., Barham G.J., Bischof R.J., Nash A., Meusen E.N.T.;
 RT "Isolation and characterization of a novel eosinophil-specific
 RT galectin."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF443208; AAL37895.1; -
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 1.
 DR SMART; SM00276; GLECT; 1.
 DR PROSITE; PS00309; GALACTIN; 1.
 SQ SEQUENCE 162 AA; 18194 MW; 0372ELF0CD5CFB4C CRC64;

Query Match 15.2%; Score 206.5; DB 6; Length 162;
 Best Local Similarity 37.1%; Pred. No. 3.9e-10;
 Matches 53; Conservative 25; Mismatches 48; Indels 17; Gaps 8;

QY 115 PY-NLPLP-----GGVPRMLITLITGV-KPNANRIALDFQRG---NDVAFHF-PREN 161
 DB 21 PYNNLPFGVVRMLITLITGVKPNANRIALDFQ-RGNDVAFHF-PRENNNRRIYVGN 171
 QY 162 ENNRRIYVGNLKNNGREEROSVFPESGKPEKIQVLEVPDHFKVAVNDA-HLYNNR 220
 DB 81 GSG--YVVCNTMQLGWWGPEKKMQPPQKSLFEICFYVDSSESFVTVNGSIFLDYAR 138
 QY 221 VKKLNIISKLIGSDIDLTSASY 243
 DB 139 L-PFEQVNAISIGGCVHVSYSIF 160

RESULT 33

Q82HF3 PRELIMINARY; PRT; 903 AA.
 ID Q82HF3
 AC Q82HF3;


```

QY 11 LSGSGNPNQGWPGMAGNQPAQ-----AGYPPASVYEGYPCQAPP---GAYPGQAPP 60
DB 1 MSYGPYP-PTGYPPFPYPPAGGESSPTTAGYPP-----YSGFPFPMGGAPP-PAPSG 52
QY 61 AYHGAAGAPP---GAPAGVYPPG--PPSGGAPAS-----SGQSPAPG-AYA 101
DB 53 GYPPGA-GGYPPAGGYPPAGGYPPGALSPPGGPPAPGGQGGAPPGAGGFGSGYPPQPPASXYG 111
QY 102 TSPY-----GAPAGPLIVPNLPLPGVVP-----RMLITLITVKENANRINALDFQR 149
DB 112 GGPAGVPPVPGFPGGMPGSGY-----PGGAPYPPSPASMTGCTGTTILASN---FDANK 164
QY 150 GNDVAFHFRFNNRNVVNCITKLDNNRGREERGVSPPESGK 193
DB 165 DAELIRKAMKGFDTDEQALVDVVSNNRNDROQIKAKAFKTYGK 208

RESULT 39
ID 0804G3 PRELIMINARY; PRT; 485 AA.
AC 0804G3;
DB 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DB Annexin 11b.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Father S.A., Olson E.S., Halpern M.E.;
RT "The Zebrafish Annexin Gene Family.";
RL Genome Res. 0:0-0(2003).
DR EMBL: AY178802; AAC020276.1; -.
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR GO: GO:0005544; F:calcium-dependent phospholipid binding; IEA.
DR InterPro: IPR001464; Annexin.
DR Pfam: PF00191; annexin; 4.
DR Pfam: PF02162; XYPPX; 9.
DR PRINTS: PR00196; ANNEXIN.
DR PRODOM: PD000143; Annexin; 4.
DR SMART: SM00335; ANX; 4.
DR PROSITE: PS00223; ANNEXIN; 4.
SQ SEQUENCE 485 AA; 51461 MW; 59D4E05C4FA253C CRC64;

Query Match 14.2%; Score 192.5; DB 13; Length 485;
Best Local Similarity 39.7%; Pred. No. 2.4e-08;
Matches 52; Conservative 12; Mismatches 34; Indels 33; Gaps 9;

QY 21 GMPGAMGNQPAAGAGY--PGASVYEGYPCQAPP---GAYPGQAPPAGYPPHAPAPAPG 76
DB 5 GYPPAGSGYPPASGTYQPPAAGYPPQAGYPPQAGYPPQ--PGAFFPPQGAFF--PQPG 60
QY 77 VYP---GPPSGPAGYPPS-----GQPSA-PGAYATGP-----YGAPAGPL 112
DB 61 AFPAGAGYPPAGGYPPAPAGGGPPPGAGYPPAGAPGAYPNMPPAGGNGHGFAGPAGCM 120
QY 113 IVPP-NLPLPG 122
DB 121 PQGYGVDPAG 131

RESULT 40
ID 071391 PRELIMINARY; PRT; 483 AA.
AC 071391;
DB 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
SQ SEQUENCE FROM N.A.

```

```

DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22386257; PubMed=12477932;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner I., Scheffer C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Burow K.H., Shemen C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.U., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.T.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalski U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maitra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RC TISSUE=Kidney;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC053208; AAH53208.1; -.
KW Hypothetical protein.
SQ SEQUENCE 483 AA; 51425 MW; E4F7F527E30FEDD7 CRC64;

Query Match 14.1%; Score 191; DB 13; Length 483;
Best Local Similarity 38.4%; Pred. No. 3.3e-08;
Matches 56; Conservative 6; Mismatches 48; Indels 36; Gaps 10;

QY 11 LSGSGNPNQGWPGMAGNQPAAGYPPAS--YGYPCQAPP---GAYPGQAPPAGYPPHAP 66
DB 1 MSYGPYPQSGSYPPQGGYPPQAGYPPAGGYPPQPPQAGYPPQ--PGAFFPPQ 58
QY 67 GAYGAPAPGYPPGPPSG-----PG-----AYPSGQSPAPG-AYATGPY 105
DB 59 GAFPG--QPGYPSVPSGGMGAPIGLNLPPHGNASNIQMGPGGMPPQNGMYPPGSP- 115
QY 106 GAPAGPLVPP-NLPLP---GVVPP 127
DB 116 --PPGQMPSTYPPNIPATPSGSPSYR 139

RESULT 41
ID 039115 PRELIMINARY; PRT; 177 AA.
AC 039115;
DB 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Glycine and proline rich protein (Genomic DNA, chromosome 5, P1
DE clone:MFCl9) (AT5g45350/MFC19_1) (Hypothetical protein).
GN GPRP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=cv. E.COLI LB392;
 RX MEDLINE=96189273; PubMed=8605310;
 RA Marty I., Monfort A., Stiefel V., Ludevid D., Delsen M.,
 RA Puigdomenech P.;
 RT "Molecular characterization of the gene coding for GPRP, a class of
 RT proteins rich in glycine and proline interacting with membranes in
 RT Arabidopsis thaliana";
 RL Plant Mol. Biol. 50:625-636(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. E.COLI LB392;
 RA Puigdomenech P.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=9397451; PubMed=10470850;
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 RT Sequence features of the regions of 1,011,550 bp covered by seventeen
 RT P1 and TAC clones.";
 RL DNA Res. 6:183-195(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Barth J.,
 RA Bowser U., Carrinchi P., Dale U.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida U., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai U., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker U.R.;
 RT "Arabidopsis cDNA clones.";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Barth J.,
 RA Bowser U., Carrinchi P., Chang E., Dale U.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida U., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai U., Lam B., Lee J.M., Lin J., Miranda M., Natusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker U.R.;
 RT "Arabidopsis ORF clones.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Haas B.U., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldman K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldman K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP EMBL: X8315; CA59059.1; -
 DR EMBL: AB018113; BAB09163.1; -
 DR EMBL: AY052250; AAK97720.1; -
 DR EMBL: AY060514; AAL31127.1; -
 DR EMBL: AY086217; AAM64295.1; -
 DR PIR: S65780; S65780.
 DR InterPro: IPR006031; XYPX.
 DR Pfam: PF02162; XYPX; 11.
 KW Hypothetical protein.
 SQ SEQUENCE 177 AA; 17830 MW; 3EF8094DCAD13F92 CRC64;

Query Match 14.0%; Score 190.5; DB 10; Length 177;
 Best Local Similarity 44.6%; Pred. No. 1,1e-08;

Matches 50; Conservative 4; Mismatches 41; Indels 17; Gaps 6;
 QY 13 GSGNPNPGWGMGNQAG--AGYFGASVP--GYGQAPPGAYPPQA--PPGAYHGA 66
 DB 2 GGDNDNDND--KGFHGYPPAGYPPPGAYPPAGYPPQGYG--PPGAYPPAGYPPGAYPP 58
 QY 67 GAYGAPAPGAYPPGPPS-----GPGAYSSGQSPAPGAYATGPGAYAPAG 110
 DB 59 GGYPPAPGAYPPAPGAYPPAGYPPAGHGYPPAGYPPAHSHGAGIGMIMG 110
 RESULT 42
 ID 044796 PRELIMINARY; PRT; 329 AA.
 AC 044796;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN C50D2.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCB1_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sammons L., Wohldmann P., Bauer C.;
 RT "The sequence of C. elegans cosmid C50D2.";
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF040642; AAB94952.1; -
 DR PIR: T32783; T32783.
 DR WormPep; C50D2.4; CE08889.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002486; Col cuticle_N.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF01484; Col cuticle_N; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 329 AA; 32080 MW; 730C7B6101A6BD9A CRC64;
 Query Match 14.0%; Score 189.5; DB 5; Length 329;
 Best Local Similarity 43.4%; Pred. No. 2.8e-08;
 Matches 53; Conservative 1; Mismatches 45; Indels 23; Gaps 7;
 QY 19 PQGPGAMGNO--PAGAGGYFGASYPYRG-----QAPPGAYPPGAPPGA-----YHGAPGA 68
 DB 187 PTGPPGPPGKQGGKGNHGPGA--PGTPGNAIPGPPGPPGPGGAGUGLPGNPPGAPGV 244
 QY 69 -----YGGAPAPGAYPPGPPGPGAYSSGQSPAPGAYATGPGAYAPAGPLI--VPYNLPL 120
 DB 245 PGQVIVPGLTPGA--GPPGPPGPIGPPGQGRAPSSSQPGPGPGGDDIGDAPGNPGS 301
 QY 121 PG 122
 DB 302 PG 303
 RESULT 43
 Q804G4

```

ID 0804G4 PRELIMINARY; PRT; 526 AA.
AC 0804G4;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Annelin 11a.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Father S.A., Olson E.S., Halpern M.E.;
RT "The Zebrafish Annelin Gene Family.";
RL Genome Res. 0:0-0(2003).
DR EMBL; AV178801; AAC20275.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IEA.
DR InterPro; IPR001464; Annelin.
DR InterPro; IPR006031; XYPX.
DR Pfam; PF00191; annelin; 4.
DR PRINTS; PR00196; ANNELIN.
DR ProDom; PD000143; Annelin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNELIN; 4.
SQ SEQUENCE 526 AA; 55732 MW; B54010208D8F79F1 CRC64;

Query Match 14.0%; Score 189.5; DB 13; Length 526;
Best Local Similarity 36.8%; Pred. No. 4.9e-08;
Matches 56; Conservative 10; Mismatches 47; Indels 39; Gaps 9;

QY 11 LSGSGNPNQCMFGAMGNQPAAGAGYPGAS--YPCYPGQAPP--GAYPGQAPPGAYNHAP 66
DB 1 MSYGYPPQSGYPPQGGYPPQPGAXYPPAGGYPPQPMYPPQAGYPPQ--PGAYPPQF 58
QY 67 GAYGAPAPGYPPGPPSG-----PGAYPS-----SCQPSAPGAYATGPY--- 105
DB 59 GAFPG--QPGQYPSVPSGSGWGPATGLDNLPNPFGNASNTQGMANQFADGGAFPNPSMFS 116
QY 106 -----GAPGAPLIV-----PYNL-PLPGSVPR 127
DB 117 GGYGPGQPGGPPANSPNOPIGMYPQGGGMPQ 148

RESULT 44
Q45904 PRELIMINARY; PRT; 285 AA.
ID 045904;
AC 045904;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE W09H1.6b protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z82081; CAB04960.1; -.
DR PIR; T26325; T26325.
DR HSSP; P17931; 1A3X.

```

```

DR WormPep; W09H1.6b; CE16577.
DR InterPro; IPR008985; Cona_1like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALACTIN; 2.
SQ SEQUENCE 285 AA; 32683 MW; BCB3D635A5848767 CRC64;

Query Match 13.9%; Score 188.5; DB 5; Length 285;
Best Local Similarity 36.4%; Pred. No. 2.8e-08;
Matches 47; Conservative 26; Mismatches 49; Indels 7; Gaps 5;

QY 114 VPYNLPLPGVVRMLITLITGVKNANRILDPQRGN-DVAHF-PPRENNRRIVCN 171
DB 156 VPYESGLANGLPVGSKLIVGVTEKKARFHNILRKNGDISFHNPRDEKH--VIN 212
QY 172 TKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKNVN-DAHQVNHRYVKKLEISKL 230
DB 213 SLAANWGNFEREGNPFKEGVGFDLVIQNEEYAFQVFNNGERYISFAHRADP-HDIAGL 271
QY 231 GISGDIDL 239
DB 272 QISGDIELS 280

RESULT 45
Q9NV1 PRELIMINARY; PRT; 277 AA.
ID Q9NV1;
AC Q9NV1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Galectin.
GN HCO-GAL-3A.
OS Haemonchus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20183859; PubMed=10717307;
RA Greenhalgh C.J.; Loukas A.; Donald D.; Nikolaou S.; Newton S.B.;
RT "A family of galectins from haemonchus contortus.";
RL Mol. Biochem. Parasitol. 107:117-121(2000).
DR EMBL; AF105668; AAF63405.1; -.
DR HSSP; P17931; 1A3X.
DR InterPro; IPR008985; Cona_1like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALACTIN; 2.
SQ SEQUENCE 277 AA; 31627 MW; 4B4CA7A97982BB82 CRC64;

Query Match 13.8%; Score 187.5; DB 5; Length 277;
Best Local Similarity 37.2%; Pred. No. 3.3e-09;
Matches 48; Conservative 20; Mismatches 54; Indels 7; Gaps 5;

QY 114 VPYNLPLPGVVRMLITLITGVKNANRILDPQRGN-DVAHF-PPRENNRRIVCN 171
DB 148 VPYESGLANGLPVGSKLIVGVTEKKARFHNILRKNGDISFHNPRDEKH--KAVIRN 204
QY 172 TKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKNVN-DAHQVNHRYVKKLEISKL 230
DB 205 ALAANWGNFEREGNPFKEGVGFDLVIQNEEYAFQVFNNGERYISFAHRADP-QDPNDISGL 263
QY 231 GISGDIDL 239
DB 264 QISGDIEL 272

RESULT 46
Q9NV0 PRELIMINARY; PRT; 283 AA.
ID Q9NV0;

```

```

AC Q9NVL0:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Galectin.
GN HCO-GAL-3B.
OS Haemochus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchus.
OX NCB1_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2018359; PubMed=10717307;
RA Greenhalgh C.J.; Loukas A.; Donald D.; Nikolaou S.; Newton S.E.;
RT "A family of galectins from haemonchus contortus."
RL Mol. Biochem. Parasitol. 107:117-121(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jiangsu.
RA Li C.H.; Wei X.F.; Li X.R.;
RT "Cloning and expression of galectin gene from Haemonchus contortus."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF105969; AAF63406.1; -
DR EMBL; AY553330; AAP5997.1; -
DR HSSP; P17931; 1A3K.
DR InterPro; IPR008985; Cona_like_lect_gl.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 283 AA; 32544 MW; F1B2E0DB346BF134 CRC64;

Query Match 13.8%; Score 187.5; DB 5; Length 283;
Best Local Similarity 37.2%; Pred. No. 3.4e-08;
Matches 48; Conservative 20; Mismatches 54; Indels 7; Gaps 5;

QY 114 VRYNPLPGGVPRMLITLGVKPNARIALDFQRGN--VAHFH-PFENNRRVIVCN 171
DB 154 VYESGSIASGFPIDKTLIFGVKEKAKFPNNILRRNDIALHFRPDE--KAVIN 210
QY 172 TRLDNNGREEROSVPEFESGKPEFKIQVLVEPDHFVAVN-DAHLQYNRVKVKLEISFL 230
DB 211 ALAANWGNWEEEREGKKEFKVGFDAIGNVAFAQIFVNGERTSFAR-CQPMDISGL 269
QY 231 GISGIDLT 239
DB 270 QIQGDIET 278

RESULT 47
Q20922 PRELIMINARY; PRT; 313 AA.
AC Q20922;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE F57B1.3 Protein.
GN F57B1.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCB1_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.A.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).

```

```

DR EMBL; Z78064; CAB01509.1; -
DR PIR; J22828; T22828.
DR WormPep; F57B1.3; CE11290.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
SQ SEQUENCE 313 AA; 29902 MW; DSA2C9DF8DCFI9E CRC64;

Query Match 13.8%; Score 187.5; DB 5; Length 313;
Best Local Similarity 43.9%; Pred. No. 3.9e-06;
Matches 50; Conservative 2; Mismatches 41; Indels 21; Gaps 6;

QY 14 SGNPNQMPGAMWGNOPAGAGYV-----GASYPYGPQAPPYGAAPPYHGAAPGA- 68
DB 172 AGPSPSAPQKG--PSGAGVGGQTGASLPPGPAGPSPSGQPGSNMAGAPGAP 229
QY 69 -----YPGAPAPGVYPPSPGYP-----SSGQSPAPAYATGPYGAAPG 111
DB 230 GQVVDVPGTPGPAGPPGPP-GPAGAPGQPGSGSGQPGPG--PQDAGAPGAP 280

RESULT 48
Q96Q57 PRELIMINARY; PRT; 300 AA.
AC Q96Q57;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Beta-galactoside-binding protein galectin-8.
GN LGALS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Merlier J.; Moisan S.; Potworowski E.F.; St-Pierre Y.;
RT "New galectin-8 isoform specific to NHL."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF037304; AAK69827.1; -
DR InterPro; IPR008985; Cona_like_lect_gl.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 300 AA; 33739 MW; DC26B39C314DE947 CRC64;

Query Match 13.8%; Score 187; DB 4; Length 300;
Best Local Similarity 38.1%; Pred. No. 4.1e-08;
Matches 51; Conservative 16; Mismatches 53; Indels 14; Gaps 7;

QY 111 PLIVPNPLPGGVPRMLITLGVKPNARIALDFQRGN--VAHFH-PFENNRR 166
DB 169 PPARRLNTMGGGV-----VKGFNANASFNVDLAGSKRIALHNPRL--NIK 220
QY 167 VIVCNTKLDNNGREERQ--SVPEFESGKPEFKIQVLVEPDHFVAVNDAR-LQYNRVKYL 224
DB 221 APVRNSFLQESNGEERNTSPFESGWFEMITVDVVEFVAVGVASHLEKHFEL 280
QY 225 NEISKIGISGIDL 238
DB 281 SSIDTLEINGIDL 294

RESULT 49
Q9M0L8 PRELIMINARY; PRT; 179 AA.
AC Q9M0L8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein (Atg19200).

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:27:21 ; Search time 59 seconds

(without alignments)
1178.079 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357
Sequence: 1 MADNPSLHDAUSGSGNPNPQ.....ISKIGISQDIDLTASASYMT 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : A_Geneseq_29Jan04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1357	100.0	246	2 AAW61955	AAW61955 Human gal
2	1357	100.0	246	6 ABU04683	ABU04683 Human exp
3	1315	96.9	250	5 AAU97819	AAU97819 Human cel
4	1315	96.9	250	6 AAG79835	AAG79835 Galectin-
5	1315	96.9	250	6 AAG79836	AAG79836 Human exp
6	1315	96.9	250	6 ABU04685	ABU04685 Human exp
7	1315	96.9	250	6 ABU04677	ABU04677 Human exp
8	1315	96.9	250	6 ABU04676	ABU04676 Human exp
9	1315	96.9	250	6 ADA11009	ADA11009 Human CDN
10	1310	96.5	249	8 ADE77133	ADE77133 Human pro
11	1310	96.5	249	7 ABU04686	ABU04686 Human exp
12	1310	96.5	249	7 ADE56588	ADE56588 Human pro
13	1305	96.2	250	2 AAW71218	AAW71218 Beta-D-ga
14	1305	96.2	250	5 AAU97818	AAU97818 Human cel
15	1305	96.2	250	6 ABU04684	ABU04684 Human exp
16	1305	96.2	250	6 ABU04674	ABU04674 Human exp
17	1305	96.2	250	6 ADC53844	ADC53844 Human gal
18	1297	95.6	250	5 AAU97820	AAU97820 Human cel
19	1297	95.6	250	5 AAU97816	AAU97816 Human cel
20	1297	95.6	250	5 AAU97817	AAU97817 Human cel
21	1297	95.6	250	6 ABU89751	ABU89751 Protein d
22	1297	95.6	250	6 ABU04680	ABU04680 Human exp
23	1297	95.6	250	6 ABU04687	ABU04687 Human exp
24	1297	95.6	250	6 ABU04678	ABU04678 Human exp
25	1297	95.6	250	6 ABU04679	ABU04679 Human exp

26	1297	95.6	250	6 ABU04675	ABU04675 Human exp
27	1297	95.6	277	4 AAG75013	AAG75013 Human col
28	1297	95.6	277	6 ABU04681	ABU04681 Human exp
29	1287	94.8	250	6 ABR56317	ABR56317 Human gal
30	1194	88.0	248	2 AAR12532	AAR12532 Human Mac
31	1194	88.0	248	6 ABU04682	ABU04682 Human exp
32	1082	79.7	264	2 AAR12531	AAR12531 Mac2.16 e
33	1082	79.7	264	7 ADC53841	ADC53841 Mouse gal
34	1082	79.7	278	2 AAR13338	AAR13338 Mac-2 pro
35	1079	78.5	262	5 AAU97821	AAU97821 Rat cell
36	1079	78.5	262	7 ADC53847	ADC53847 Rat cell
37	1078	79.4	264	2 AAW71219	AAW71219 Beta-D-ga
38	1078	79.4	264	6 ABU04782	ABU04782 Human exp
39	1075	79.2	262	2 AAR42300	AAR42300 IGE bindi
40	1074	79.1	261	7 ADE56586	ADE56586 Rat Prote
41	1064.5	78.4	263	5 AAU97822	AAU97822 Mouse cel
42	1060	78.1	258	6 AAW61954	AAW61954 Rat galec
43	723	53.3	143	6 AAG79833	AAG79833 N-termina
44	723	53.3	144	6 AAG79834	AAG79834 N-termina
45	710	52.3	141	6 ABR56316	ABR56316 N-termina
46	593	43.7	138	1 AAB60534	AAB60534 C-termina
47	314.5	23.2	341	4 AAB97171	AAB97171 Rainbow t
48	299	22.0	322	7 ADC53842	ADC53842 Mouse gal
49	289	21.3	355	2 AAY06997	AAY06997 Galectin-
50	289	21.3	355	2 AAW85664	AAW85664 Galectin-
51	289	21.3	355	5 ABB77852	ABB77852 Amino aci
52	289	21.3	355	7 ADC53845	ADC53845 Human gal
53	289	21.3	355	7 ADE62929	ADE62929 Human pro
54	289	21.3	355	7 ADD48101	ADD48101 Human pro
55	289	21.3	378	4 AAE13847	AAE13847 Human lun
56	289	21.3	378	7 ADE67474	ADE67474 Human lun
57	289	21.3	378	7 ADE88001	ADE88001 Human lun
58	283	20.9	323	3 AAY56802	AAY56802 Human eos
59	283	20.9	323	5 ABB77853	ABB77853 Amino aci
60	283	20.9	323	5 ABB61494	ABB61494 Human NF-
61	282	20.8	329	5 AAU97036	AAU97036 Human bla
62	277.5	20.4	322	7 ABU63650	ABU63650 Rat urate
63	276.5	20.4	354	7 ADC53848	ADC53848 Rat galec
64	276.5	20.4	354	7 ADE62927	ADE62927 Rat Prote
65	276.5	20.4	354	7 ADD48099	ADD48099 Rat Prote

ALIGNMENTS

RESULT 1	
ID	AAW61955 standard; protein; 246 AA.
XX	AC AAW61955;
XX	AC AAW61955;
XX	DT 18-SEP-1998 (first entry)
DE	Human galectin amino acid sequence.
DE	Human galectin; diabetes-mediating protein; insulin; DMP; diabetes;
KW	drug screening assay.
XX	OS Homo sapiens.
XX	WO9820124-A2.
XX	PD 14-MAY-1998.
XX	EF 24-OCT-1997; 97WO-IB001627.
XX	PR 05-OCT-1996; 96US-0029324P.
XX	PR 05-NOV-1996; 96US-0030088P.
XX	PR 05-NOV-1996; 96US-0030186P.
XX	PR 18-JUL-1997; 97US-00897098.
XX	(LARS/) MOSE LARSEN P.
PA	(FEYS/) FEY S J.

(NERU//) NERUP J.
PA (KARL//) KARLESEN A E.
PA (CHR//) BOERRE CHRISTENSEN U.
PA (POCI//) POCIOT F.
PA (ANDR//) ANDERSEN H U.
PI Mose Larsen P, Fey SJ, Nerup J, Karlseu AE, Bjerre Christensen U;
PI Pociot F, Andersen HU;
XX
DR WFI; 1998-286940/25.
XX
PT Identification of diabetes-mediating protein(s) - by transplanting
PT insulin-secreting cells into host at risk of developing diabetes and
PT analysing protein expression in transplanted cells.
XX
PS Example 5; Fig 5; 15app; English.
XX
CC This represents the amino acid sequence of murine mortalin. This is a
CC diabetes-mediating protective protein used in the method of invention.
CC The invention provides methods for in vivo identification of diabetes-
CC mediating protein (DMP) by transplanting insulin-secreting cells into
CC host at risk of developing diabetes and analysing protein expression in
CC transplanted cells. The DMPs are useful in drug screening assays for
CC identifying compounds capable of modulating the development of diabetes,
CC useful as therapeutic agents for the treatment or prevention of diabetes,
CC and useful as targets of therapeutic agents capable of preventing or
CC ameliorating diabetes by modulating the expression of the DMP. Changes in
CC the expression of specific DMPs is diagnostically useful as indicator of
CC the development of diabetes
SO Sequence 246 AA;
XX
Query Match 100.0%; Score 1357; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 7.9e-100;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MADNPSLHDAISGSNPDPQGMPPGAMGNOPAGAGSPGASYPGYRGAPPGAYPQGAPP 60
DB 1 MADNSLHDALSGSNPNPDQGMPPGAMGNOPAGAGSPGASYPGYRGAPPGAYPQGAPP 60
OY 61 AVHGAPGAYPGAFAVGVPVGPSPGGGAAYPSSSQPSAPGAYATPGYAGAPGLIVEYNPL 120
DB 61 AVHGAPGAYPGAFAVGVPVGPSPGGGAAYPSSSQPSAPGAYATPGYAGAPGLIVEYNPL 120
OY 121 PGGVPRMLITLTIGTVKKPANRIALDFQRGDVAHFPPFENNRRVIYCNTKLNNMGR 180
DB 121 PGGVPRMLITLTIGTVKKPANRIALDFQRGDVAHFPPFENNRRVIYCNTKLNNMGR 180
OY 181 EBRQSVFPESGSKPKRIQVLVEPDHPFKYAVNDNAHIQYNNRVVKLEISKLGISGIDITS 240
DB 181 EBRQSVFPESGSKPKRIQVLVEPDHPFKYAVNDNAHIQYNNRVVKLEISKLGISGIDITS 240

OY 241 ASYTWI 246
DB 241 ASYTWI 246

RESULT 2
ABU04683
ID ABU04683 standard; protein; 246 AA.
XX
AC ABU04683;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1349.
XX
KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MEC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX

OS	Homo sapiens.
XX	WO200278524-A2.
XX	10-OCT-2002.
PD	
XX	28-MAR-2002; 2002WO-US009671.
PF	
XX	28-MAR-2001; 2001US-0279495P.
PR	21-MAY-2001; 2001US-0292544P.
PR	08-AUG-2001; 2001US-0310801P.
PR	01-OCT-2001; 2001US-0326370P.
PR	04-DEC-2001; 2001US-0336780P.
PR	20-FEB-2002; 2002US-0358985P.
PA	(ZYCO-) ZYCOS INC.
XX	
PI	Chicz RM, Tomlinson AJ, Urban RG;
XX	WPI; 2003-040607/03.
XX	
XX	New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT	cytoskeletal proteins, receptors or transcription factors), useful for
PT	treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT	leukemia.
XX	
PS	Example 2; SEQ ID NO 1349; 134pp; English.
XX	
CC	The invention describes a purified polypeptide, which comprises a
CC	fragment of a kinase, phosphatase, protease, protease inhibitor,
CC	transporter, cytoskeletal protein, receptor or transcription factor. The
CC	polypeptide is useful as an immunogenic composition for eliciting in a
CC	mammal an immunogenic response directed against any of the purified
CC	polypeptide. The purified polypeptide, or the antibody that binds to this
CC	polypeptide, is useful for treating cancer. The polypeptide is also
CC	useful for identifying compounds that binds to a naturally processed
CC	class I or class II MHC-binding polypeptide. The polypeptides and
CC	polynucleotides are particularly useful for treating or preventing
CC	lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC	leukemia or leukemia. These are also useful for screening agents for
CC	treating the above mentioned diseases. This sequence represents an
CC	expressed protein tag (SPR) isolated from human tissue for translational
CC	profiling. Note: This sequence does not appear in the printed
CC	specification but was obtained in electronic format directly from WIPO at
XX	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 246 AA;
Query Match	100.0%; Score 1357; DB 6; Length 246;
Best Local Similarity	100.0%; Pred. No. 7.9e-100;
Matches 246; Conservative	0; Mismatches 0; Indels 0
1	MADNDSIHDALSSGNPNPQGMFGAMGNQPAAGVYPASVYGGVGGAPRGAYPGAPPG 60
2	
3	1 MADNDSIHDALSSGNPNPQGMFGAMGNQPAAGVYPASVYGGVGGAPRGAYPGAPPG 60
4	
5	61 AYHGAPGAPGAPGAPGAPGVGPGPSGCGAVPSSGQPSAPGAYATGYPGAPAPPLIVYNLEL 120
6	
7	61 AYHGAPGAPGAPGAPGAPGVGPGPSGCGAVPSSGQPSAPGAYATGYPGAPAPPLIVYNLEL 120
8	
9	61 AYHGAPGAPGAPGAPGAPGVGPGPSGCGAVPSSGQPSAPGAYATGYPGAPAPPLIVYNLEL 120
10	
11	121 PGGVVPRMLITLIGTVKPNANFIADPFGKGNVAHFHPRFNNRNVIVCNTKLDNNMR 180
12	
13	121 PGGVVPRMLITLIGTVKPNANFIADPFGKGNVAHFHPRFNNRNVIVCNTKLDNNMR 180
14	
15	121 PGGVVPRMLITLIGTVKPNANFIADPFGKGNVAHFHPRFNNRNVIVCNTKLDNNMR 180
16	
17	181 EERGQVPEPSGKPPKIVLVEPDHPKAVVNDAAHQYNRKYKKNLEISKLGISGIDILTS 240
18	
19	181 EERGQVPEPSGKPPKIVLVEPDHPKAVVNDAAHQYNRKYKKNLEISKLGISGIDILTS 240
20	
21	241 ASYTM 246
22	
23	241 ASYTM 246
24	
25	241 ASYTM 246
26	
27	241 ASYTM 246
28	
29	241 ASYTM 246
30	
31	241 ASYTM 246
32	
33	241 ASYTM 246
34	
35	241 ASYTM 246
36	
37	241 ASYTM 246
38	
39	241 ASYTM 246
40	
41	241 ASYTM 246
42	
43	241 ASYTM 246
44	
45	241 ASYTM 246
46	
47	241 ASYTM 246
48	
49	241 ASYTM 246
50	
51	241 ASYTM 246
52	
53	241 ASYTM 246
54	
55	241 ASYTM 246
56	
57	241 ASYTM 246
58	
59	241 ASYTM 246
60	
61	241 ASYTM 246
62	
63	241 ASYTM 246
64	
65	241 ASYTM 246
66	
67	241 ASYTM 246
68	
69	241 ASYTM 246
70	
71	241 ASYTM 246
72	
73	241 ASYTM 246
74	
75	241 ASYTM 246
76	
77	241 ASYTM 246
78	
79	241 ASYTM 246
80	
81	241 ASYTM 246
82	
83	241 ASYTM 246
84	
85	241 ASYTM 246
86	
87	241 ASYTM 246
88	
89	241 ASYTM 246
90	
91	241 ASYTM 246
92	
93	241 ASYTM 246
94	
95	241 ASYTM 246
96	
97	241 ASYTM 246
98	
99	241 ASYTM 246
100	
101	241 ASYTM 246
102	
103	241 ASYTM 246
104	
105	241 ASYTM 246
106	

```
RESULT 3
AAU97819
ID .AAU97819 standard; protein; 250 AA.
XX
XX
AC AAU97819;
XX
XX
DT 27-AUG-2002 (first entry)
XX
DE Human cell membrane anchor protein galectin-3 #4.
XX
XX
KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;
KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
KW polycystic kidney; post-angioplasty restenosis; cytosolic;
KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
KW vasotropic; hepatotropic; human.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200229031-A2.
XX
XX
PD 11-APR-2002.
XX
XX
PF 01-OCT-2001; 2001WO-IL000918.
XX
XX
PR 04-OCT-2000; 2000US-0237858P.
XX
XX
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
XX
PI Kloog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
XX
XX
DR WPI; 2002-435333/46.
XX
XX
PT Identifying anchor proteins that bind Ras protein, by producing complexes
PT of Ras and cell membrane proteins in the presence and absence of a Ras
PT antagonist and identifying a complex disrupted by the Ras antagonist.
XX
XX
PS Disclosure; Page 13; 62pp; English.
XX
XX
CC The invention describes a method of identifying cell membrane anchor
CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
CC where one mixture has a Ras antagonist. A cross linking agent is added,
CC and complexes between Ras protein and other proteins are produced. The
CC complexes are then separated and the proteins binding to Ras are
CC identified. The invention also describes a method useful for identifying
CC drug candidates that inhibit aberrant Ras activity. An antisense compound
CC comprising at least one phosphorothioate-modified nucleotide is useful
CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
CC compound into a patient exhibiting this problem. The method is also
CC useful for identifying anchor proteins for the farnesylated isoforms of H
CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
CC for treating diseases characterized by uncontrolled mitosis, including
CC cancers and various non-malignancies such as autoimmune disease (e.g.
CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty
CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
CC protein that binds an isoform of Ras
XX
XX
SQ Sequence 250 AA;
XX
XX
Query Match 96.9%; Score 1315; DB 5; Length 250;
Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
```

```
QY 1 MADNFSLHDALSSGNNPQGMWPGAMGNQPAAGAGYPGASYPG-YPCQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSSGNNPQGMWPGAMGNQPAAGAGYPGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHGAPGAYPGAPGAYPGGPGSPGAYPSSGQPSAPGAY-ATGPYGAAPGLIVNYL 118
DB 61 GAYHGAPGAYPGAPGAYPGGPGSPGAYPSSGQPSAPGAYPATGPYGAAPGLIVNYL 120
```

```
QY 119 PLPGGVPRMLITLIGTVKPNANRIALDPQRGDVAFHE-PRFNENNRVIVCNKTKLDNN 177
DB 121 PLPGGVPRMLITLIGTVKPNANRIALDPQRGDVAFHEPRFNENNRVIVCNKTKLDNN 180
QY 178 WGREERQSVFPESGKPFKIQVLVEPDHFKVANDAH-LQYNHRVYKLEISKLGISGDI 236
DB 181 WGREERQSVFPESGKPFKIQVLVEPDHFKVANDAHILQYNHRVYKLEISKLGISGDI 240
QY 237 DITSASYMTI 246
DB 241 DITSASYMTI 250
RESULT 4
AAG79835
ID AAG79835 standard; protein; 250 AA.
XX
XX
AC AAG79835;
XX
XX
DT 16-APR-2003 (first entry)
XX
XX
DE Galectin-3.
XX
XX
KW Galectin-3; nuclear localisation; BHL domain; Bcl-2; metastasis;
KW N-terminally truncated; polyethylene glycol; tumour; cancer;
KW carbohydrate binding; multimerisation; gene therapy.
XX
XX
OS Homo sapiens.
XX
XX
PN WO2002100343-A2.
XX
XX
PD 19-DEC-2002.
XX
XX
PF 10-JUN-2002; 2002WO-US018478.
XX
XX
PR 08-JUN-2001; 2001US-0296970P.
XX
XX
PA (MAND-) MANDALMED INC.
XX
XX
PI Jarvis GA, John CM, Lefler H;
XX
XX
DR WPI; 2003-183915/18.
XX
XX
PT Novel N-terminally truncated galectin-3 derivatized with polyethylene
PT glycol, useful for treating tumor.
XX
XX
PS Disclosure; Page 23; 88pp; English.
XX
XX
CC The sequences given in AAG79833-35 represent recombinant human galectin-
CC 3's. Galectins bind lactose and the human protein is composed of 250
CC amino acids with a molecular weight of approx. 31000. N-terminally
CC truncated galectin-3 derivatized with at least one molecule of
CC polyethylene glycol is useful for treating a tumour in a patient and for
CC reducing tumour size, and for treating cancer or for preventing
CC metastasis. It is also useful for inhibiting carbohydrate binding and
CC multimerisation of galectin-3, and for inhibiting tumour growth and
CC metastasis in vivo. DNA encoding the truncated galectin-3 is useful in
CC gene therapy
XX
XX
SQ Sequence 250 AA;
XX
XX
Query Match 96.9%; Score 1315; DB 6; Length 250;
Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
```

```
QY 1 MADNFSLHDALSSGNNPQGMWPGAMGNQPAAGAGYPGASYPG-YPCQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSSGNNPQGMWPGAMGNQPAAGAGYPGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHGAPGAYPGAPGAYPGGPGSPGAYPSSGQPSAPGAY-ATGPYGAAPGLIVNYL 118
DB 61 GAYHGAPGAYPGAPGAYPGGPGSPGAYPSSGQPSAPGAYPATGPYGAAPGLIVNYL 120
```

```

QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFORGNDVAEHE-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFORGNDVAEHEFNNRRVIVCNTKLDNN 180
QY 178 WGERERQSVFPESGKPKIQLVEPDHFKVAVNDAAH-LQYNHRVKKLINEISKLGISGDI 236
DB 181 WGERERQSVFPESGKPKIQLVEPDHFKVAVNDAAHLLQYNHRVKKLINEISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 5
AAG79836
ID AAG79836 standard; protein; 250 AA.
XX
AC AAG79836;
XX
DT 16-APR-2003 (first entry)
XX
DE Galectin-3.
XX
KM Galectin-3; nuclear localisation; BHL domain; Bcl-2; metastasis;
KM N-terminally truncated; polyethylene glycol; tumour; cancer;
KM carbohydrate binding; multimerisation; gene therapy.
XX
OS Homo sapiens.
XX
PN MO2002100343-A2.
XX
PD 19-DEC-2002.
XX
PF 10-JUN-2002; 2002WO-US018478.
XX
PR 08-JUN-2001; 2001US-0296970P.
XX
PI (MAND-) MANDALMED INC.
XX
PI Jarvis GA, John CM, Jelfler H;
XX
DR WPI; 2003-183915/18.
XX
DR N-PSDB; ABA00863.
XX
PT Novel N-terminally truncated galectin-3 derivatized with polyethylene
PT glycol, useful for treating tumor.
XX
PS Disclosure; Page 59; 88pp; English.
XX
XX
CC This sequence shows full length human galectin-3. Galectins bind lactose
CC and the human protein is composed of 250 amino acids with a molecular
CC weight of approx. 31000. N-terminally truncated galectin-3 derivatised
CC with at least one molecule of polyethylene glycol is useful for treating
CC a tumour in a patient and for reducing tumour size, and for treating
CC cancer or for preventing metastasis. It is also useful for inhibiting
CC carbohydrate binding and multimerisation of galectin-3, and for
CC inhibiting tumour growth and metastasis in vivo. DNA encoding the
CC truncated galectin-3 is useful in gene therapy
XX
SQ Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;
Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
QY 1 MADNFSIHDALSGSGNPNPGMPCAMGNQAPAGAGYFGASYPG-YFGQAPPGAYPGQAPP 59
DB 1 MADNFSIHDALSGSGNPNPGMPCAMGNQAPAGAGYFGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHAGAGAYPGADAPGVYGPSPGPGAYPSSGQPSAPGAY-ATGPIGAPAGPIITVYNL 118
DB 61 GAYHAGAGAYPGADAPGVYGPSPGPGAYPSSGQPSAPGAYPATGPIYAGAPAGPIITVYNL 120

```

```

QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFORGNDVAEHE-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFORGNDVAEHEFNNRRVIVCNTKLDNN 180
QY 178 WGERERQSVFPESGKPKIQLVEPDHFKVAVNDAAH-LQYNHRVKKLINEISKLGISGDI 236
DB 181 WGERERQSVFPESGKPKIQLVEPDHFKVAVNDAAHLLQYNHRVKKLINEISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 6
ABU04685
ID ABU04685 standard; protein; 250 AA.
XX
AC ABU04685;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1351.
XX
KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KM protease; protease inhibitor; transporter; cytoskeletal protein;
KM receptor; transcription factor; cancer; MHC;
KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS Homo sapiens.
XX
PN MO200278524-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-US009671.
XX
PR 28-MAR-2001; 2001US-0279495P.
XX
PR 21-MAY-2001; 2001US-0292544P.
XX
PR 08-AUG-2001; 2001US-0310801P.
XX
PR 01-OCT-2001; 2001US-0326370P.
XX
PR 04-DEC-2001; 2001US-0336780P.
XX
PR 20-FEB-2002; 2002US-0358985P.
XX
PA (ZYCO-) ZYCOS INC.
XX
PI Chicx RM, Tomlinson AJ, Urban RG;
XX
DR WPI; 2003-040607/03.
XX
XX
CC New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
CC cytoskeletal proteins, receptors or transcription factors), useful for
CC treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
CC leukemia.
XX
PT Example 2; SEQ ID NO 1351; 134pp; English.
XX
PS
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed

```

CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;
 Best Local Similarity 98.4%; Pred. No. 1.7e-96;
 Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MAFNPSLHDAISGSGNPNPQGMWPGAMGNOPAGAGYPPGASYPG-YFGQAPPGAYPGQAP 59
 DB 1 MAFNPSLHDAISGSGNPNPQGMWPGAMGNOPAGAGYPPGASYPGAYPGQAPPGAYPGQAP 60
 QY 60 GAYHGAAPGAYPGAPGAYPPGSGGAGAPSSGQPSAPGAY-ATGPGYAPAGPLIYPYNL 118
 DB 61 GAYHGAAPGAYPGAPGAYPPGSGGAGAYPPSSGQPSAPGAYPATGPGYAPAGPLIYPYNL 120
 QY 119 PLPGGVVPRMLITILGTVPKPNANRIALDFQGNDAVAFHFNPRNENNRVIVCNTKLDNN 177
 DB 121 PLPGGVVPRMLITILGTVPKPNANRIALDFQGNDAVAFHFNPRNENNRVIVCNTKLDNN 180
 QY 178 WGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNEISKLGISDI 236
 DB 181 WGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAAHLLQYNHRVKKLNEISKLGISDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 7
 ABU04677
 ID ABU04677 standard; protein; 250 AA.
 XX
 AC ABU04677;
 XX
 DT 29-JAN-2003 (first entry)
 XX

DE Human expressed protein tag (EPT) #1343.
 XX

KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KM protease; protease inhibitor; transporter; cytoskeletal protein;
 KM receptor; transcription factor; cancer; MHC;
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX

OS Homo sapiens.
 XX

PN WO200278524-A2.
 XX

PD 10-OCT-2002.
 XX

PF 28-MAR-2002; 2002WO-US009671.
 XX

PR 28-MAR-2001; 2001US-0279495P.
 XX

PR 21-MAY-2001; 2001US-0292544P.
 XX

PR 08-AUG-2001; 2001US-0310801P.
 XX

PR 01-OCT-2001; 2001US-0326370P.
 XX

PR 04-DEC-2001; 2001US-0336780P.
 XX

PR 20-FEB-2002; 2002US-0358985P.
 XX

PA (ZYCO-) ZYCOS INC.
 XX

PI Chiciz RM, Tomlinson AJ, Urban RG;
 XX

DR WPI; 2003-040607/03.
 XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukaemia.
 XX
 PS Example 2; SEQ ID NO 1343; 134P; English.

XX The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;
 Best Local Similarity 98.4%; Pred. No. 1.7e-96;
 Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MAFNPSLHDAISGSGNPNPQGMWPGAMGNOPAGAGYPPGASYPG-YFGQAPPGAYPGQAP 59
 DB 1 MAFNPSLHDAISGSGNPNPQGMWPGAMGNOPAGAGYPPGASYPGAYPGQAPPGAYPGQAP 60
 QY 60 GAYHGAAPGAYPGAPGAYPPGSGGAGAPSSGQPSAPGAY-ATGPGYAPAGPLIYPYNL 118
 DB 61 GAYHGAAPGAYPGAPGAYPPGSGGAGAYPPSSGQPSAPGAYPATGPGYAPAGPLIYPYNL 120
 QY 119 PLPGGVVPRMLITILGTVPKPNANRIALDFQGNDAVAFHFNPRNENNRVIVCNTKLDNN 177
 DB 121 PLPGGVVPRMLITILGTVPKPNANRIALDFQGNDAVAFHFNPRNENNRVIVCNTKLDNN 180
 QY 178 WGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNEISKLGISDI 236
 DB 181 WGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAAHLLQYNHRVKKLNEISKLGISDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 8
 ABU04676
 ID ABU04676 standard; protein; 250 AA.
 XX
 AC ABU04676;
 XX
 DT 29-JAN-2003 (first entry)
 XX

DE Human expressed protein tag (EPT) #1342.
 XX

KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KM protease; protease inhibitor; transporter; cytoskeletal protein;
 KM receptor; transcription factor; cancer; MHC;
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX

OS Homo sapiens.
 XX

PN WO200278524-A2.
 XX

PD 10-OCT-2002.
 XX

PF 28-MAR-2002; 2002WO-US009671.
 XX

PR 28-MAR-2001; 2001US-0279495P.
 XX

PR 21-MAY-2001; 2001US-0292544P.
 XX

PR 08-AUG-2001; 2001US-0310801P.
 XX

PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCO INC.

XX Chicz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.

PS Example 2; SEQ ID NO 1342; 1342P; English.

XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor.
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (BPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;

Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

```

QY 1 MNDNPSLHDLSSGSGNPNQGMWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MANDFSLHDLSSGSGNPNQGMWGNOPAGAGYPGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHGAPGAYPGAPAGYVPPSPGAYPSSGQPSAPGAY-ATGPGAPAGELIVPNYL 118
DB 61 GAYHGAPGAYPGAPAGYVPPSPGAYPSSGQPSAPGAYPATGPGAYGAPGLIVPNYL 120
QY 119 PLPGVAVPRLITLIGTVKPNANRIALDFQGNDAVFHF-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGVAVPRLITLIGTVKPNANRIALDFQGNDAVFHFPRFENNRRVIVCNTKLDNN 180
QY 178 WGERESQVPPFSGRPFKIQVLEVPDHFKAIVNDAH-LOYNHRVKKLNISKLGISGDI 236
DB 181 WGERESQVPPFSGRPFKIQVLEVPDHFKAIVNDAHLLQYNHRVKKLNISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

```

RESULT 9

ADAL1009 standard; protein; 250 AA.

XX ADAL1009;

XX 06-NOV-2003 (first entry)

XX Human cDNA differentially expressed in colon cancer #85 product.

XX differential expression; colon cancer; cancer; human.

XX Homo sapiens.

XX US2002160382-A1.

XX 31-OCT-2002.

XX 11-OCT-2001; 2001US-00981353.

XX 11-OCT-2000; 2000US-0239841P.

XX (LASEK/) LASEK A W.

XX (JONES/) JONES D A.

XX Lasek AW, Jones DA;

XX WPI; 2003-265756/26.

XX N-PSDB; ADAL1008.

XX New combination comprising cDNAs that are differentially expressed in
PT colon disorder, useful for diagnosing, treating, staging or monitoring
PT treatment for colon cancers.

PS Example 14; SEQ ID NO 127; 2312P; English.

XX The invention relates to a combination comprising cDNAs that are
CC differentially expressed in colon disorder. The methods and compositions
CC of the present invention are useful for diagnosing, treating, staging or
CC monitoring treatment for colon cancer. They are also useful in high
CC throughput methods for using cDNAs to detect differential expression of
CC nucleic acids in a sample, screening molecules or compounds to identify a
CC ligand which specifically binds a cDNA and using a protein to screen
CC molecules or compounds to identify at least one ligand which specifically
CC binds the protein. The present sequence represents the amino acid
CC sequence of a human cDNA differentially expressed in colon cancer
CC protein.

XX Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;

Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

```

QY 1 MNDNPSLHDLSSGSGNPNQGMWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MANDFSLHDLSSGSGNPNQGMWGNOPAGAGYPGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHGAPGAYPGAPAGYVPPSPGAYPSSGQPSAPGAY-ATGPGAPAGELIVPNYL 118
DB 61 GAYHGAPGAYPGAPAGYVPPSPGAYPSSGQPSAPGAYPATGPGAYGAPGLIVPNYL 120
QY 119 PLPGVAVPRLITLIGTVKPNANRIALDFQGNDAVFHF-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGVAVPRLITLIGTVKPNANRIALDFQGNDAVFHFPRFENNRRVIVCNTKLDNN 180
QY 178 WGERESQVPPFSGRPFKIQVLEVPDHFKAIVNDAH-LOYNHRVKKLNISKLGISGDI 236
DB 181 WGERESQVPPFSGRPFKIQVLEVPDHFKAIVNDAHLLQYNHRVKKLNISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

```

RESULT 10

ADE77133 standard; protein; 250 AA.

XX ADE77133;

XX 29-JAN-2004 (first entry)

XX Human protein expressed in a liver disorder #77.

XX human; liver disorder; hyperlipidaemia; hypertension; type II diabetes;
KM tumour; liver; inflammatory disorder; immune response disorder;
KM high-throughput screening; differential gene expression; gene therapy.
OS Homo sapiens.
XX US2003108871-A1.
PN 12-JUN-2003.
PD 30-JUL-2001; 2001US-00919039.
XX 28-JUL-2000; 2000US-0222113P.
PR (KASE/) KASER M R.
XX Kaser MR;
PI MPI: 2004-031227/03.
DR N-PSDB; ADE77132.
XX Composition comprising several cDNAs that are differentially expressed in
PT treated human C3A liver cell cultures, useful for treating liver
PT disorders.
XX Claim 1; SEQ ID NO 298, 41pp; English.
PS The invention relates to a composition comprising several cDNAs that are
CC differentially expressed in a liver disorder. The composition is useful
CC for treating liver disorder such as hyperlipidaemia, hypertension, type
CC II diabetes, tumours of the liver and disorders of the inflammatory and
CC immune response. The composition is useful for a high-throughput method
CC of screening several molecules or compounds to identify a ligand which
CC specifically binds a cDNA. A protein encoded by the cDNA is useful for a
CC high-throughput method for using a protein to screen several molecules or
CC compounds to identify at least one ligand which specifically binds the
CC protein which involves combining the protein encoded by the cDNA with
CC several of molecules or compounds under conditions to allow specific
CC binding, and detecting specific binding between the protein and a
CC molecule or compound, therefore identifying a ligand which specifically
CC binds the protein. The composition is useful for detecting and
CC quantifying differential gene expression, can be used in gene therapy, to
CC formulate prognosis and to design a treatment regimen and to monitor the
CC efficacy of treatment. The present sequence represents the amino acid
CC sequence of a protein encoded by a cDNA differentially expressed in a
CC liver disorder.
SQ Sequence 250 AA;
Query Match 96.9%; Score 1315; DB 8; Length 250;
Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPG-YPGQAPPGAYPGGAPP 59
DB 1 MADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPGAYPGQAPPGAYPGGAPP 60
QY 60 GAYHGAAGAYPGAPAPGVYPGPSGPGAYSSGQPSAPGAY-ATGPGAGAGAPLIVVYNL 118
DB 61 GAYHGAAGAYPGAPAPGVYPGPSGPGAYSSGQPSAPGAYPATGPGAGAGPLIVVYNL 120
QY 119 PLTGGVPRMLITLITLGTAKFNANRIALDPORGNDVAFH-PRFNNENRRTIVCCTKLDNN 177
DB 121 PLTGGVPRMLITLITLGTAKFNANRIALDPORGNDVAFHPRFNNENRRTIVCCTKLDNN 180
QY 178 WGREROSVFPFSSGKPFKIOVLVEPDHFKVAVNDAA-LOYNRRVKKLNEISKLGISGDI 236
DB 181 WGREROSVFPFSSGKPFKIOVLVEPDHFKVAVNDAAILOYNRRVKKLNEISKLGISGDI 240
QY 237 DLTSAASYTMI 246
DB 241 DLTSAASYTMI 250

RESULT 11
ABU04686
ID ABU04686 standard; protein; 249 AA.
XX AC ABU04686;
XX 29-JAN-2003 (first entry)
DE Human expressed protein tag (EPT) #1352.
XX Translational profiling: expressed protein tag; EPT; kinase; phosphatase;
KM protease; protease inhibitor; transporter; cytoskeletal protein;
KM receptor; transcription factor; cancer; MHC;
KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX Homo sapiens.
OS WO200278524-A2.
PN 10-OCT-2002.
XX 28-MAR-2002; 2002WO-US009671.
PF 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985P.
XX (ZYCO-) ZYCOG INC.
PI Chiciz RM, Tomlinson AJ, Urban RG;
DR MPI: 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
PS Example 2; SEQ ID NO 1352; 134pp; English.
XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC ftp.wipo.int/pub/published_pcl_sequences
XX Sequence 249 AA;
Query Match 96.5%; Score 1310; DB 6; Length 249;
Best Local Similarity 98.4%; Pred. No. 4.3e-96;
Matches 245; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 2 ADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPG-YPGQAPPGAYPGGAPP 60
DB 1 ADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPGAYPGQAPPGAYPGGAPP 60

QY 61 AYHGAPGAYPGAPAGVYPPGSPGAGAYSSGQSPAPGAY-ATGPGYAPAGPLIYVYMLP 119
DB 61 AYHGAPGAYPGAPAGVYPPGSPGAGAYSSGQSPAPGAYPATGPGYAPAGPLIYVYMLP 120
QY 120 LPGAIVPRLITLIGTVKPNANRILADPQKNDVAFHF-PPENNNRRVIVCNTKLDNNW 178
DB 121 LPGAIVPRLITLIGTVKPNANRILADPQKNDVAFHF-PPENNNRRVIVCNTKLDNNW 180
QY 179 GREERQSVFPFESGKPFKIQVIVPEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGDI 237
DB 181 GREERQSVFPFESGKPFKIQVIVPEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGDI 240
QY 238 LTSASYTMI 246
DB 241 LTSASYTMI 249
RESULT 12
ADE56588 standard; protein; 249 AA.
XX ADE56588;
AC ADE56588;
XX 29-JAN-2004 (first entry)
DT Human Protein P17931, SEQ ID NO 2442.
XX Human Protein P17931, SEQ ID NO 2442.
DE Human; pain; neuronal tissue; gene therapy;
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
KW spinal segmental nerve injury; SN1; Chung.
XX Homo sapiens.
OS Homo sapiens.
XX WO2003016475-A2.
PN 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
PF 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-033347P.
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX Woolf C, D'ureo D, Belfort K, Costrigan M;
PI WPI; 2003-268312/26.
XX GENBANK; P17931.
DR New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
PT Claim 1; Page; 1017p; English.
XX The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
XX claimed are a vector comprising the novel polynucleotide, a host cell
XX comprising the vector, a method for identifying a nucleotide sequence
XX which is differentially regulated in an animal subjected to pain and a
XX kit to perform the method, an array, a method for identifying an agent
XX that increases or decreases the expression of the polynucleotide sequence
XX that is differentially expressed in neuronal tissue of a first animal
XX subjected to pain, a method for identifying a compound which regulates
XX the expression of a polynucleotide sequence which is differentially
XX expressed in an animal subjected to pain, a method for identifying a
XX compound that regulates the activity of one or more of the
XX polynucleotides, a method for producing a pharmaceutical composition, a
XX method for identifying a compound or small molecule that regulates the
XX activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_ptc_sequences.
XX Sequence 249 AA;
SQ
Query Match 96.5%; Score 1310; DB 7; Length 249;
Best Local Similarity 98.4%; Pred. No. 4.3e-96;
Matches 245; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
QY 2 ADNFSIHDALSGSGNPNQGMFGAWGNQPAAGGYPGASVYRG-YPGQAPPGAYPQQAAPG 60
DB 1 ADNFSIHDALSGSGNPNQGMFGAWGNQPAAGGYPGASVYPGAYPQQAAPG 60
QY 61 AYHGAPGAYPGAPAGVYPPGSPGAGAYSSGQSPAPGAY-ATGPGYAPAGPLIYVYMLP 119
DB 61 AYHGAPGAYPGAPAGVYPPGSPGAGAYSSGQSPAPGAYPATGPGYAPAGPLIYVYMLP 120
QY 120 LPGAIVPRLITLIGTVKPNANRILADPQKNDVAFHF-PPENNNRRVIVCNTKLDNNW 178
DB 121 LPGAIVPRLITLIGTVKPNANRILADPQKNDVAFHF-PPENNNRRVIVCNTKLDNNW 180
QY 179 GREERQSVFPFESGKPFKIQVIVPEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGDI 237
DB 181 GREERQSVFPFESGKPFKIQVIVPEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGDI 240
QY 238 LTSASYTMI 246
DB 241 LTSASYTMI 249
RESULT 13
AAW71218 standard; protein; 250 AA.
XX AAW71218;
AC AAW71218;
XX 30-OCT-1998 (first entry)
DT Beta-D-galactoside-binding protein designated L-31-gal-lectin.
XX Beta-D-galactoside-binding protein; L-31-gal-lectin;
DE Beta-D-galactoside-binding protein; L-31-gal-lectin;
KW metastatic potential; antibody.
XX Homo sapiens.
OS Homo sapiens.
XX US5801002-A.
PN 01-SEP-1998.
PD 22-NOV-1995; 95US-00562311.
XX 06-JAN-1989; 89US-00294249.
PR 05-APR-1991; 91US-00681242.
XX 26-JAN-1994; 94US-00188225.
XX (KARM-) KARMANOS CANCER INST BARBARA ANN.
PA Raz A;
PI WPI; 1998-494766/42.
XX N-PSDB; AAV54735.
XX Test for metastatic potential of cell sample - by measuring binding of
PT antibody to L-31-gal-lectin on cell surface.

XX Claim 1; Fig 6A-B; 24pp; English.
 PS
 CC The present sequence represents a beta-D-galactoside-binding protein
 CC designated L-31-gal-lectin. The sequence is derived from clone 1. The
 CC specification describes a method for testing a cell sample for metastatic
 CC potential. The method comprises contacting the sample with a labelled
 CC antibody that binds to endogenous cell-surface L-31-gal-lectin, removing
 CC unbound antibody, and determining the amount of bound antibody as a
 CC measure of L-31-gal-lectin expression, where the metastatic potential
 CC increases as the level of L-31-gal-lectin expression increases. The
 CC antibody is produced by immunisation with a L-31-gal-lectin protein
 XX
 SQ Sequence 250 AA;
 Query Match 96.2%; Score 1305; DB 2; Length 250;
 Best Local Similarity 97.6%; Pred. No. 1.1e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;
 QY 1 MADNFSLHDALSSGNPNPQGWPGAMGNQPAAGGYPGASYPG-YPGQAPPGAYPGQAPP 59
 DB 1 MADNFSLHDALSSGNPNPQGWPGAMGNQPAAGGYPGASYPGAYPGQAPPAYPGQAPP 60
 QY 60 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPGSSGQPSAGAPAY-ATGPGYAPAPGLIYPPYL 118
 DB 61 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPGSSGQPSAGAPAYPATGPGYAPAPGLIYPPYL 120
 QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFORGNDVAFFH-PRFNNRNRYIVCNTKLDNN 177
 DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFORGNDVAFFHPRFNNRNRYIVCNTKLDNN 180
 QY 178 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDAH-LQYNHRVKKINEISKLGISGDI 236
 DB 181 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDAHLLQYNHRVKKINEISKLGISGDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250
 RESULT 14
 ID AAU97818 standard; protein; 250 AA.
 AC AAU97818;
 DT 27-AUG-2002 (first entry)
 DE Human cell membrane anchor protein galectin-3 #3.
 KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;
 KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
 KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
 KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
 KW polycystic kidney; post-angioplasty restenosis; cyostatic;
 KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
 KW vasotropic; hepatotropic; human.
 OS Homo sapiens.
 WO200229031-A2.
 PD 11-APR-2002.
 PF 01-OCT-2001; 2001WO-IL000918.
 PR 04-OCT-2000; 2000US-0237858P.
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
 PI KIoog Y, Haklai R, Paz A, El Ad-Sfadia G, Barlan E;
 DR WPI; 2002-435333/46.

XX Identifying anchor proteins that bind Ras protein, by producing complexes
 PT of Ras and cell membrane proteins in the presence and absence of a Ras
 PT antagonist and identifying a complex disrupted by the Ras antagonist.
 XX
 PS Disclosure; Page 13; 62pp; English.
 CC The invention describes a method of identifying cell membrane anchor
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
 CC where one mixture has a Ras antagonist. A cross linking agent is added,
 CC and complexes between Ras protein and other proteins are produced. The
 CC complexes are then separated and the proteins binding to Ras are
 CC identified. The invention also describes a method useful for identifying
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound
 CC comprising at least one phosphorothioate-modified nucleotide is useful
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
 CC compound into a patient exhibiting this problem. The method is also
 CC useful for identifying anchor proteins for the farnesylated isoforms of H
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
 CC for treating diseases characterised by uncontrolled mitosis, including
 CC cancers and various non-malignancies such as autoimmune disease (e.g.
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
 CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
 CC protein that binds an isoform of Ras
 XX
 SQ Sequence 250 AA;
 Query Match 96.2%; Score 1305; DB 5; Length 250;
 Best Local Similarity 98.0%; Pred. No. 1.1e-95;
 Matches 245; Conservative 0; Mismatches 1; Indels 4; Gaps 4;
 QY 1 MADNFSLHDALSSGNPNPQGWPGAMGNQPAAGGYPGASYPG-YPGQAPPGAYPGQAPP 59
 DB 1 MADNFSLHDALSSGNPNPQGWPGAMGNQPAAGGYPGASYPGAYPGQAPPAYPGQAPP 60
 QY 60 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPGSSGQPSAGAPAY-ATGPGYAPAPGLIYPPYL 118
 DB 61 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPGSSGQPSAGAPAYPATGPGYAPAPGLIYPPYL 120
 QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFORGNDVAFFH-PRFNNRNRYIVCNTKLDNN 177
 DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFORGNDVAFFHPRFNNRNRYIVCNTKLDNN 180
 QY 178 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDAH-LQYNHRVKKINEISKLGISGDI 236
 DB 181 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDAHLLQYNHRVKKINEISKLGISGDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250
 RESULT 15
 ID ABU04684 standard; protein; 250 AA.
 AC ABU04684;
 DT 29-JAN-2003 (first entry)
 DE Human expressed protein tag (EPT) #1350.
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 OS Homo sapiens.
 WO200278524-A2.

```
XX 10-OCT-2002.
PD 28-MAR-2002; 2002WO-US009671.
XX 28-MAR-2002; 2002WO-US009671.
XX 28-MAR-2001; 2001US-0279495P.
XX 21-MAY-2001; 2001US-0292544P.
XX 08-AUG-2001; 2001US-0310801P.
XX 01-OCT-2001; 2001US-0326370P.
XX 04-DEC-2001; 2001US-0336780P.
XX 20-FEB-2002; 2002US-0358985P.
XX (ZYCO-) ZYCOS INC.
XX Chicz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX Example 2; SEQ ID NO 1350; 134pp; English.
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor. The
XX polypeptide is useful as an immunogenic composition for eliciting in a
XX mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to this
XX polypeptide, is useful for treating cancer. The polypeptide is also
XX useful for identifying compounds that binds to a naturally processed
XX class I or class II MHC-binding polypeptide. The polypeptides and
XX polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukaemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling. Note: This sequence does not appear in the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 250 AA;
SQ
Query Match 96.2%; Score 1305; DB 6; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.1e-95;
Matches 24; Conservative 0; Mismatches 2; Indels 4; Gaps 4;
QY 1 MADNFSIHDALSSGNNPQGMPCGAWGNOPAGAGYGCASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADNFSIHDALSSGNNPQGMPCGAWGNOPAGAGYGCASYPGAYPGQAPPAYPGQAPP 60
QY 60 GAYHGAAGAYPGAPAPGVYPPSGPGAYPSGPGASPGAY-ATGPGYAPAGPLIVPNYL 118
DB 61 GAYHGAAGAYPGAPAPGVYPPSGPGAYPSGPGASPGAYATGPGYAPAGPLIVPNYL 120
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNVVAHFENFRNNRRTVVCNTKLDNN 177
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNVVAHFENFRNNRRTVVCNTKLDNN 180
QY 178 WGRERQSVFPPESGKPKLOVVEPDHFKYAVNDAH-LOVNHVVKKLNEISKIGISDI 236
DB 181 WGRERQSVFPPESGKPKHVLVBDHFKA VandahlLQTNHVKKLNEIRKIGISDI 240
QY 237 DLTASATMTI 246
DB 241 DLTASATMTI 250
```

RESULT 16
ABU04674
ID ABU04674 standard; protein; 250 AA.

```
XX ABU04674;
AC 29-JAN-2003 (first entry)
XX Human expressed protein tag (EPT) #1340.
XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
XX protease; protease inhibitor; transporter; cytoskeletal protein;
XX receptor; transcription factor; cancer; MHC;
XX major histocompatibility complex; myeloma; colon cancer; gastric cancer;
XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX Homo sapiens.
XX WO200278524-A2.
XX 10-OCT-2002.
XX 28-MAR-2002; 2002WO-US009671.
XX 28-MAR-2001; 2001US-0279495P.
XX 21-MAY-2001; 2001US-0292544P.
XX 08-AUG-2001; 2001US-0310801P.
XX 01-OCT-2001; 2001US-0326370P.
XX 04-DEC-2001; 2001US-0336780P.
XX 20-FEB-2002; 2002US-0358985P.
XX (ZYCO-) ZYCOS INC.
XX Chicz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX Example 2; SEQ ID NO 1340; 134pp; English.
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor. The
XX polypeptide is useful as an immunogenic composition for eliciting in a
XX mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to this
XX polypeptide, is useful for treating cancer. The polypeptide is also
XX useful for identifying compounds that binds to a naturally processed
XX class I or class II MHC-binding polypeptide. The polypeptides and
XX polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukaemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling. Note: This sequence does not appear in the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 250 AA;
SQ
Query Match 96.2%; Score 1305; DB 6; Length 250;
Best Local Similarity 98.0%; Pred. No. 1.1e-95;
Matches 24; Conservative 0; Mismatches 1; Indels 4; Gaps 4;
QY 1 MADNFSIHDALSSGNNPQGMPCGAWGNOPAGAGYGCASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADNFSIHDALSSGNNPQGMPCGAWGNOPAGAGYGCASYPGAYPGQAPPAYPGQAPP 60
QY 60 GAYHGAAGAYPGAPAPGVYPPSGPGAYPSGPGASPGAY-ATGPGYAPAGPLIVPNYL 118
DB 61 GAYHGAAGAYPGAPAPGVYPPSGPGAYPSGPGASPGAYATGPGYAPAGPLIVPNYL 120
```

QY 119 PLPGVYVPRMLITITIGVKNANRIALDPORGNDVAFHF-PRENNRRVIVCNTKLDNN 177
 DB 121 PLPGVYVPRMLITITIGVKNANRIALDPORGNDVAFHFENPRNENRRVIVCNTKLDNN 180
 QY 178 WGEERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNRVKKLEISLGISSGI 236
 DB 181 WGEERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLLQYNRVKKLEISLGISSGI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 17
 ADCS3844
 ID ADCS3844 standard; protein; 250 AA.
 AC ADCS3844;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human galectin 3 protein.
 XX
 KW galectin-3; nephritis; glomerular nephritis; antiinflammatory;
 KW glomerular infiltration; apoptosis; human.
 XX
 OS Homo sapiens.
 EN JP2002322082-A.
 XX
 PD 08-NOV-2002.
 XX
 PF 26-APR-2001; 2001JP-00129200.
 XX
 PR 26-APR-2001; 2001JP-00129200.
 XX
 PA (PROT-) PROTEGENE KK.
 XX
 DR WPI; 2003-367092/35.
 XX
 PT Agents for prevention and treatment of nephritis, comprise galectin-1,
 PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration
 PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8
 PT positive cells.
 XX
 PS Disclosure; SEQ ID NO 5; 31pp; Japanese.
 XX
 CC This invention relates to the use of novel mammal derived galectin-1
 CC (G1), -3 (G3) and -9 (G9) proteins as effective ingredients for
 CC prevention and treatment of nephritis. The invention discloses agents for
 CC prevention and treatment of nephritis, particularly glomerular nephritis
 CC and may have antiinflammatory activities. The method of the invention
 CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and
 CC apoptosis of CD8 positive cells. The method and sequences of the
 CC invention may be used for prevention and treatment of nephritis,
 CC particularly glomerular nephritis including inhibition of glomerular
 CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8
 CC positive cells. The present sequence represents the human galectin 3
 CC protein of the invention.
 XX
 SQ Sequence 250 AA;

Query Match 96.2%; Score 1305; DB 7; Length 250;
 Best Local Similarity 98.0%; Pred. No. 1.1e-95;
 Matches 245; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

QY 1 MADNFSJHDLALSGSGNPNPGWPGAWGNOPAGAGYVPGASYPG-YPGQAPPGAYPGQAPP 59
 DB 1 MANNFSJHDLALSGSGNPNPGWPGAWGNOPAGAGYVPGASYPGAYPGQAPP 60
 QY 60 GAYHAGAPGAYGAPAPGVPPGPPSGPAGYSSQGPSAPGAY-ATGPGAGAPGLIIVYNL 118
 DB 61 GAYHAGAPGAYGAPAPGVPPGPPSGPAGYSSQGPSAPGAYPATGPGAGAPGLIIVYNL 120

QY 119 PLPGVYVPRMLITITIGVKNANRIALDPORGNDVAFHF-PRENNRRVIVCNTKLDNN 177
 DB 121 PLPGVYVPRMLITITIGVKNANRIALDPORGNDVAFHFENPRNENRRVIVCNTKLDNN 180
 QY 178 WGEERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNRVKKLEISLGISSGI 236
 DB 181 WGEERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLLQYNRVKKLEISLGISSGI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 18
 AAU97820
 ID AAU97820 standard; protein; 250 AA.
 AC AAU97820;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE Human cell membrane anchor protein galectin-3 #5.
 XX
 KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;
 KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
 KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
 KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
 KW polycystic kidney; post-angioplasty restenosis; cyostatic;
 KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
 KW vasotropic; hepatotropic; human.
 XX
 OS Homo sapiens.
 EN WO200229031-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 01-OCT-2001; 2001WO-IL000918.
 XX
 PR 04-OCT-2000; 2000US-0237858P.
 XX
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
 XX
 PI Klooog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
 XX
 DR WPI; 2002-435333/46.
 XX
 PT Identifying anchor proteins that bind Ras protein, by producing complexes
 PT of Ras and cell membrane proteins in the presence and absence of a Ras
 PT antagonist and identifying a complex disrupted by the Ras antagonist.
 XX
 PS Disclosure; Page 14; 62pp; English.

The invention describes a method of identifying cell membrane anchor
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
 CC where one mixture has a Ras antagonist. A cross linking agent is added,
 CC and complexes between Ras protein and other proteins are produced. The
 CC complexes are then separated and the proteins binding to Ras are
 CC identified. The invention also describes a method useful for identifying
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound
 CC comprising at least one phosphorothioate-modified nucleotide is useful
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
 CC compound into a patient exhibiting this problem. The method is also
 CC useful for identifying anchor proteins for the farnesylated isoforms of H
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
 CC for treating diseases characterized by uncontrolled mitosis, including
 CC cancers and various non-malignancies such as autoimmune disease (e.g.
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
 CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
 CC protein that binds an isoform of Ras

XX SQ Sequence 250 AA;
 Query Match 95.6%; Score 1297; DB 5; Length 250;
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNLSLHDALSGSGNPNQGMWGNQAPAGAGYPCASYPG-YPGQAPGAYGQAP 59
 DB 1 MADNLSLHDALSGSGNPNQGMWGNQAPAGAGYPCASYPGAYPGQAPGAYGQAP 60
 QY 60 GAYHAPGAYPGAPAPGAYPPGSPGAYPSSGQSPAPGAY-ATGPGYAPAGPLIVPNL 118
 DB 61 GAYPAPGAYPGAPAPGAYPPGSPGAYPSSGQSPATGAYPATGPGYAPAGPLIVPNL 120
 QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQRNDVAFHF-PRFNNRRVIVCNTKLDNN 177
 DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQRNDVAFHFPRFNNRRVIVCNTKLDNN 180
 QY 178 WGRERQSYFPESGKPEFKIQVLVEPDHKVAVNDAH-LQYHRYKLNLSKLGISGDI 236
 DB 181 WGRERQSYFPESGKPEFKIQVLVEPDHKVAVNDAHLLQYHRYKLNLSKLGISGDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 19
 AAU97816
 ID AAU97816 standard; protein; 250 AA.
 AC AAU97816;
 DT 27-AUG-2002 (first entry)
 DE Human cell membrane anchor protein galectin-3 #1.
 XX
 XX Galectin-3; cell membrane anchor protein; Ras; antisense technology;
 XX farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
 XX cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
 XX multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
 XX polycystic kidney; post-angioplasty restenosis; cytostatic;
 XX immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
 XX vasotropic; hepatotropic; human.
 OS Homo sapiens.
 XX
 XX W0200229031-A2.
 PN 11-APR-2002.
 PD 01-OCT-2001; 2001WO-11000918.
 PF 04-OCT-2000; 2000US-0237858P.
 PR (UTRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
 PA Kluog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
 PI WPI: 2002-435333/46.
 DR N-PSDB; ABK52348.
 XX
 XX Identifying anchor proteins that bind Ras protein, by producing complexes
 PT of Ras and cell membrane proteins in the presence and absence of a Ras
 PT antagonist and identifying a complex disrupted by the Ras antagonist.
 XX
 XX Disclosure; Page 11-12; 62pp; English.
 XX
 XX The invention describes a method of identifying cell membrane anchor
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
 CC where one mixture has a Ras antagonist. A cross linking agent is added,
 CC and complexes between Ras protein and other proteins are produced. The

CC complexes are then separated and the proteins binding to Ras are
 CC identified. The invention also describes a method useful for identifying
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound
 CC comprising at least one phosphorothioate-modified nucleotide is useful
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
 CC compound into a patient exhibiting this problem. The method is also
 CC useful for identifying anchor proteins for the farnesylated isoforms of H
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
 CC for treating diseases characterised by uncontrolled mitosis, including
 CC cancers and various non-malignancies such as autoimmune disease (e.g.
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
 CC rejection, atherosclerosis, polycystic kidney and post-angioplasty
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
 CC protein that binds an isoform of Ras

XX SQ Sequence 250 AA;
 Query Match 95.6%; Score 1297; DB 5; Length 250;
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNLSLHDALSGSGNPNQGMWGNQAPAGAGYPCASYPG-YPGQAPGAYGQAP 59
 DB 1 MADNLSLHDALSGSGNPNQGMWGNQAPAGAGYPCASYPGAYPGQAPGAYGQAP 60
 QY 60 GAYHAPGAYPGAPAPGAYPPGSPGAYPSSGQSPAPGAY-ATGPGYAPAGPLIVPNL 118
 DB 61 GAYPAPGAYPGAPAPGAYPPGSPGAYPSSGQSPATGAYPATGPGYAPAGPLIVPNL 120
 QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQRNDVAFHF-PRFNNRRVIVCNTKLDNN 177
 DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQRNDVAFHFPRFNNRRVIVCNTKLDNN 180
 QY 178 WGRERQSYFPESGKPEFKIQVLVEPDHKVAVNDAH-LQYHRYKLNLSKLGISGDI 236
 DB 181 WGRERQSYFPESGKPEFKIQVLVEPDHKVAVNDAHLLQYHRYKLNLSKLGISGDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 20
 AAU97817
 ID AAU97817 standard; protein; 250 AA.
 AC AAU97817;
 DT 27-AUG-2002 (first entry)
 DE Human cell membrane anchor protein galectin-3 #2.
 XX
 XX Galectin-3; cell membrane anchor protein; Ras; antisense technology;
 XX farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
 XX cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
 XX multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
 XX polycystic kidney; post-angioplasty restenosis; cytostatic;
 XX immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
 XX vasotropic; hepatotropic; human.
 OS Homo sapiens.
 XX
 XX W0200229031-A2.
 PN 11-APR-2002.
 PD 01-OCT-2001; 2001WO-11000918.
 PF 04-OCT-2000; 2000US-0237858P.
 PR (UTRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
 PA Kluog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
 PI WPI: 2002-435333/46.
 DR N-PSDB; ABK52348.
 XX
 XX Identifying anchor proteins that bind Ras protein, by producing complexes
 PT of Ras and cell membrane proteins in the presence and absence of a Ras
 PT antagonist and identifying a complex disrupted by the Ras antagonist.
 XX
 XX Disclosure; Page 11-12; 62pp; English.
 XX
 XX The invention describes a method of identifying cell membrane anchor
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
 CC where one mixture has a Ras antagonist. A cross linking agent is added,
 CC and complexes between Ras protein and other proteins are produced. The

PI KLoog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
 XX WPI; 2002-43533/46.
 DR N-PSDB; ABK52349.
 XX
 PT Identifying anchor proteins that bind Ras protein, by producing complexes
 of Ras and cell membrane proteins in the presence and absence of a Ras
 PT antagonist and identifying a complex disrupted by the Ras antagonist.
 XX
 PS Disclosure; Page 12; 62pp; English.

XX The invention describes a method of identifying cell membrane anchor
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
 CC where one mixture has a Ras antagonist. A cross linking agent is added,
 CC and complexes between Ras protein and other proteins are produced. The
 CC complexes are then separated and the proteins binding to Ras are
 CC identified. The invention also describes a method useful for identifying
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound
 CC comprising at least one phosphorothioate-modified nucleotide is useful
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
 CC compound into a patient exhibiting this problem. The method is also
 CC useful for identifying anchor proteins for the farnesylated isoforms of H
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
 CC for treating diseases characterised by uncontrolled mitosis, including
 CC cancers and various non-malignancies such as autoimmune disease (e.g.
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
 CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
 CC protein that binds an isoform of Ras
 CC
 XX

Sequence 250 AA;

Query Match 95.6%; Score 1297; DB 5; Length 250;
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPSASYPG-YPGQAPPGAYPGQAP 59
 DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPSASYPGAYPGQAPPGAYPGQAP 60
 QY 60 GAYHAPGAYPGAPAPGAYPGPPSGGAYPSGSGSAGAY-ATGPGYAPAGPLIVPNL 118
 DB 61 GAYHAPGAYPGAPAPGAYPGPPSGGAYPSGSGSAGAYPATGPGYAPAGPLIVPNL 120
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNVAFHF-PRFNNRNRRVIVCNTKLDNN 177
 DB 121 PLPGGVPRMLITLIGTVKPNANRIALDFQGNVAFHFPRFNNRNRRVIVCNTKLDNN 180
 QY 178 WGREROSVFPFESGKPFKIQVLEPDPHFKAIVNDAH-LOYNHRVKKLNLSKIGISDI 236
 DB 181 WGREROSVFPFESGKPFKIQVLEPDPHFKAIVNDAHLOYNHRVKKLNLSKIGISDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 21
 ABU89751
 ID ABU89751 standard; protein; 250 AA.
 XX
 AC ABU89751;
 XX
 DT 10-JUL-2003 (first entry)
 XX
 XX Protein differentially expressed in cardiovascular disease #45.
 DE Cardiovascular disease; arteriosclerosis; ischaemia; angina pectoris;
 KW myocardial infarction; cardiac; antiarteriosclerotic; antianginal;
 KW gene therapy; differential gene expression.
 XX
 OS Homo sapiens.

XX
 PN WO2003031650-A2.
 XX
 PD 17-APR-2003.
 XX
 PF 02-OCT-2002; 2002WC-EP011034.
 XX
 PR 08-OCT-2001; 2001GB-00024145.
 XX
 PA (FARB) BAYER AG.
 XX

XX Munnes M, Gehrman M, Wick M, Schmitz G;
 XX WPI; 2003-403108/38.
 DR N-PSDB; ACA89924.
 XX

PT Predicting, diagnosing or prognosing a cardiovascular disease, e.g.
 PT angina, ischemia, myocardial infarction or arteriosclerosis by detection
 PT of a polynucleotide in a biological sample comprises detecting a
 PT hybridization complex.
 XX

PS Claim 3; Page 385-386; 454pp; English.

XX The invention describes a method of predicting, diagnosing or prognosing
 CC a cardiovascular disease by detection of a polynucleotide in a biological
 CC sample comprises hybridizing at least one of the polynucleotide to a
 CC nucleic acid material of a biological sample, thus forming a
 CC hybridisation complex, and detecting the hybridisation complex. The
 CC polynucleotides, polypeptides, antisense molecule, antibody and reagent
 CC are useful for preparing compositions for preventing, predicting or
 CC diagnosing, or a medication for treating a cardiovascular disease, e.g.
 CC arteriosclerosis, ischemia, angina pectoris, or myocardial infarction.
 CC This sequence represents a protein identified in the invention a being
 CC differentially expressed in individuals with cardiovascular disease
 CC
 XX

Sequence 250 AA;

Query Match 95.6%; Score 1297; DB 6; Length 250;
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPSASYPG-YPGQAPPGAYPGQAP 59
 DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPSASYPGAYPGQAPPGAYPGQAP 60
 QY 60 GAYHAPGAYPGAPAPGAYPGPPSGGAYPSGSGSAGAY-ATGPGYAPAGPLIVPNL 118
 DB 61 GAYHAPGAYPGAPAPGAYPGPPSGGAYPSGSGSAGAYPATGPGYAPAGPLIVPNL 120
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNVAFHF-PRFNNRNRRVIVCNTKLDNN 177
 DB 121 PLPGGVPRMLITLIGTVKPNANRIALDFQGNVAFHFPRFNNRNRRVIVCNTKLDNN 180
 QY 178 WGREROSVFPFESGKPFKIQVLEPDPHFKAIVNDAH-LOYNHRVKKLNLSKIGISDI 236
 DB 181 WGREROSVFPFESGKPFKIQVLEPDPHFKAIVNDAHLOYNHRVKKLNLSKIGISDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 22
 ABU04680
 ID ABU04680 standard; protein; 250 AA.
 XX
 AC ABU04680;
 XX
 DT 29-JAN-2003 (first entry)
 XX
 XX Human expressed protein tag (EPT) #1346.
 DE Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW
 XX

KM protease; protease inhibitor; transporter; cytoskeletal protein;
 KM receptor; transcription factor; cancer; MHC;
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX Homo sapiens.
 OS WO200278524-A2.
 XX PD 10-OCT-2002.
 XX PF 28-MAR-2002; 2002MO-US009671.
 XX PR 28-MAR-2001; 2001US-0279495P.
 XX PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX (ZYCO-) ZYCOS INC.
 PA Chicx RM, Tomlinson AJ, Urban RG;
 PI WPI; 2003-040607/03.
 XX DR
 XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX PS Example 2; SEQ ID NO 1346; 134pp; English.
 XX CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 250 AA;
 XX
 Query Match 95.6%; Score 1297; DB 6; Length 250;
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 237 DLTASYTMI 246
 DB 241 DLTASYTMI 250
 RESULT 23
 ID ABU04687
 ABU04687 standard; protein; 250 AA.
 AC ABU04687;
 DT 29-JAN-2003 (first entry)
 DE Human expressed protein tag (EPT) #1353.
 XX Translation profiling; expressed protein tag; EPT; kinase; phosphatase;
 KM protease; protease inhibitor; transporter; cytoskeletal protein;
 KM receptor; transcription factor; cancer; MHC;
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX Homo sapiens.
 OS WO200278524-A2.
 XX PD 10-OCT-2002.
 XX PF 28-MAR-2002; 2002MO-US009671.
 XX PR 28-MAR-2001; 2001US-0279495P.
 XX PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX (ZYCO-) ZYCOS INC.
 PA Chicx RM, Tomlinson AJ, Urban RG;
 PI WPI; 2003-040607/03.
 XX DR
 XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX PS Example 2; SEQ ID NO 1353; 134pp; English.
 XX CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 250 AA;
 XX
 Query Match 95.6%; Score 1297; DB 6; Length 250;
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

QY 1 MADNFSLHDAISGSGNPNPQGWPGWGMGNQPAAGGYPGASYPG-YRQAPPGAYPGQAPP 59
DB 1 MADNFSLHDAISGSGNPNPQGWPGWGMGNQPAAGGYPGASYPGAYPGQAPPAYPGQAPP 60
QY 60 GAYHGAAPGAYPGAPAPGVYPPGSGPGAYPSSGQPSAPGAY-ATGPGAPAGPLIYVYNL 118
DB 61 GAYPGAPGAYPGAPAPGVYPPGSGPGAYPSSGQPSATGAYPATGPGAYGAPGLIYVYNL 120
QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVHFH-PRFNNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVHFHPRFNNRRVIVCNTKLDNN 180
QY 178 WGREERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAH-LQYNHRVKKLNESKLGISDI 236
DB 181 WGREERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAHLLQYNHRVKKLNESKLGISDI 240
QY 237 DLTASASYTMI 246
DB 241 DLTASASYTMI 250

```

RESULT 24

ABU04678 ID ABU04678 standard; protein; 250 AA.

AC ABU04678;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1344.

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC;

KW major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

OS WO200278524-A2.

PN 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCO5 INC.

XX Chicz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX cytoskeletal proteins, receptors or transcription factors), useful for

XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or

XX leukemia.

XX Example 2; SEQ ID NO 1344; 134pp; English.

CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIP0 at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 250 AA;

Query Match 95.6%; Score 1297; DB 6; Length 250;
Best Local Similarity 97.6%; Pred. No. 4, 6e-95;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

QY 1 MADNFSLHDAISGSGNPNPQGWPGWGMGNQPAAGGYPGASYPG-YRQAPPGAYPGQAPP 59
DB 1 MADNFSLHDAISGSGNPNPQGWPGWGMGNQPAAGGYPGASYPGAYPGQAPPAYPGQAPP 60
QY 60 GAYHGAAPGAYPGAPAPGVYPPGSGPGAYPSSGQPSAPGAY-ATGPGAPAGPLIYVYNL 118
DB 61 GAYPGAPGAYPGAPAPGVYPPGSGPGAYPSSGQPSATGAYPATGPGAYGAPGLIYVYNL 120
QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVHFH-PRFNNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVHFHPRFNNRRVIVCNTKLDNN 180
QY 178 WGREERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAH-LQYNHRVKKLNESKLGISDI 236
DB 181 WGREERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAHLLQYNHRVKKLNESKLGISDI 240
QY 237 DLTASASYTMI 246
DB 241 DLTASASYTMI 250

```

RESULT 25

ABU04679 ID ABU04679 standard; protein; 250 AA.

AC ABU04679;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1345.

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC;

KW major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

OS WO200278524-A2.

PN 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCO5 INC.

XX Chicz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.

Example 2; SEQ ID NO 1345; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 250 AA;

XX Query Match 95.6%; Score 1297; DB 6; Length 250;

XX Best Local Similarity 97.6%; Pred. No. 4.6e-95;

XX Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

XX 1 MADNFSIHDLSSGNPNPQGWPGAWGNOPAGAGGYPGASYPG-YPGQAPPGAYPGQAPP 59

XX 1 MADNFSIHDLSSGNPNPQGWPGAWGNOPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60

XX 60 GAYHGAAPGAYPGAPAPAGVYPPSPSGGAYPSSSQSPAPGAY-ATGTYGAPAGGLIYVYNL 118

XX 61 GAYPGAPGAYPGAPAGVYPPSPSGGAYPSSSQSPATGAYPATGTGYPAGGLIYVYNL 120

XX 119 PLPGGVVPRMLITLITGVKPNANRIALDPORGNDVAFHF-PRENNRRRYIVCNTKLDNN 177

XX 121 PLPGGVVPRMLITLITGVKPNANRIALDPORGNDVAFHFPRNNRRRYIVCNTKLDNN 180

XX 178 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAH-LQYNNRVKLNKISKLGISGDI 236

XX 181 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAHLLQYNNRVKLNKISKLGISGDI 240

XX 237 DLTSASYTMI 246

XX 241 DLTSASYTMI 250

XX RESULT 26

XX AB004675 standard; protein; 250 AA.

XX AC AB004675;

XX 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #1341.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

XX protease; protease inhibitor; transporter; cytoskeletal protein;

XX receptor; transcription factor; cancer; MHC;

XX major histocompatibility complex; myeloma; colon cancer; gastric cancer;

XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

XX OS

XX WO200278524-A2.

PD 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.

XX 28-MAR-2001; 2001US-0279495P.

XX 21-MAY-2001; 2001US-0292544P.

XX 08-AUG-2001; 2001US-0310801P.

XX 01-OCT-2001; 2001US-0326370P.

XX 04-DEC-2001; 2001US-0336780P.

XX 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCOs INC.

XX Chicx RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

PT cytoskeletal proteins, receptors or transcription factors), useful for

PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or

PT leukemia.

XX Example 2; SEQ ID NO 1341; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 250 AA;

XX Query Match 95.6%; Score 1297; DB 6; Length 250;

XX Best Local Similarity 97.6%; Pred. No. 4.6e-95;

XX Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

XX 1 MADNFSIHDLSSGNPNPQGWPGAWGNOPAGAGGYPGASYPG-YPGQAPPGAYPGQAPP 59

XX 1 MADNFSIHDLSSGNPNPQGWPGAWGNOPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60

XX 60 GAYHGAAPGAYPGAPAPAGVYPPSPSGGAYPSSSQSPAPGAY-ATGTYGAPAGGLIYVYNL 118

XX 61 GAYPGAPGAYPGAPAGVYPPSPSGGAYPSSSQSPATGAYPATGTGYPAGGLIYVYNL 120

XX 119 PLPGGVVPRMLITLITGVKPNANRIALDPORGNDVAFHF-PRENNRRRYIVCNTKLDNN 177

XX 121 PLPGGVVPRMLITLITGVKPNANRIALDPORGNDVAFHFPRNNRRRYIVCNTKLDNN 180

XX 178 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAH-LQYNNRVKLNKISKLGISGDI 236

XX 181 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAHLLQYNNRVKLNKISKLGISGDI 240

XX 237 DLTSASYTMI 246

XX 241 DLTSASYTMI 250

XX RESULT 27

XX AAG75013 standard; protein; 277 AA.

XX ID AAG75013

XX

AC AAG75013;
 XX 03-SEP-2001 (first entry)
 DT
 XX
 DE Human colon cancer antigen protein SEQ ID NO:5777.
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX colorectal carcinoma; chromosome 14.
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US026524.
 XX
 PR 29-SEP-1999; 99US-0157137P.
 PR 03-NOV-1999; 99US-0163280P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI; 2001-235357/24.
 DR N-PSDB; AAH34418.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers.
 XX
 PS Claim 11; Page 7294-7295; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
 CC proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytoskeletal activity and can be used in gene therapy
 CC and vaccine production. N and P may be used in the prevention, diagnosis
 CC and treatment of diseases associated with inappropriate P expression. For
 CC example, N and P may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of P by expressing inactive proteins or to
 CC supplement the patient's own production of P. Additionally, N may be used
 CC to produce the patient's own production of P, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the proteins. N and P
 CC can be used in the prevention, diagnosis and treatment of colorectal
 CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
 CC sequences used in the exemplification of the present invention. N.B.
 CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
 CC time of publication, meaning no sequences are present for SEQ ID NO:1027
 CC to 1052, 7921 and 7922
 XX
 SQ Sequence 277 AA;
 XX
 Query Match 95.6%; Score 1297; DB 4; Length 277;
 Best Local Similarity 97.6%; Pred. No. 5.2e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;
 XX
 QY 1 MANDSELDHMLSSGPNPQGMGAMGNOGAPGAGGPGASYPG-YFGQAPPGAYPGQAPP 59
 DB 28 MANDSELDHMLSSGPNPQGMGAMGNOGAPGAGGPGASYPGAYPGQAPP 87
 QY 60 GAVHAGPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 118
 DB 88 GAVHAGPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 147
 QY 119 PLPGGVPEMLITLITGVKNNRILALDFQSGNDVAFHF-PRNENNRVIVCTLDNN 177
 DB 148 PLPGGVPEMLITLITGVKNNRILALDFQSGNDVAFHFPRNENNRVIVCTLDNN 207
 QY 178 WGREROSVFPESGPFKIQVLVEPDHKKVAVNDAA-LQYHRRVKKLEISKLGSDI 236
 DB 208 WGREROSVFPESGPFKIQVLVEPDHKKVAVNDAAHLQYHRRVKKLEISKLGSDI 267

QY 237 DLTSASYMT 246
 DB 268 DLTSASYMT 277
 XX
 RESULT 28
 ABU04681
 ID ABU04681 standard; protein; 277 AA.
 XX
 AC ABU04681;
 XX
 DT 29-JAN-2003 (first entry)
 XX
 DE Human expressed protein tag (EPT) #1347.
 XX
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200278524-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-US009671.
 XX
 PR 28-MAR-2001; 2001US-0279495P.
 PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX
 PA (ZYCO-) ZYCO INC.
 XX
 PI Chicx RM, Tomlinson AJ, Urban RG;
 XX
 DR WPI; 2003-040607/03.
 XX
 PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX
 PS Example 2; SEQ ID NO 1347; 134pp; English.
 XX
 CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 277 AA;
 XX
 Query Match 95.6%; Score 1297; DB 6; Length 277;
 Best Local Similarity 97.6%; Pred. No. 5.2e-95;
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSSGNGNPNPGWPGAMGNOGAPAGAGYFGASYPG-YPGQAPPGAYPGQAPP 59
 DB 28 MADNFSLHDALSSGNGNPNPGWPGAMGNOGAPAGAGYFGASYPGAYPGQAPP 87
 QY 60 GAYHGAAPGAPGAPAPGVYPPGPPGAPYSSSGSPAPGAY-ATGPGYGAAPGLIVRYNL 118
 DB 88 GAYHGAAPGAPGAPAPGVYPPGPPGAPYSSSGSPAPGAYPATGPGYGAAPGLIVRYNL 147
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHF-PRFNNRNRYIVCNTKLDNN 177
 DB 148 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHFNNRNRYIVCNTKLDNN 207
 QY 178 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAH-LQYNHRVKYKLINEISKIGISGDI 236
 DB 208 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAHLLQYNHRVKYKLINEISKIGISGDI 267
 QY 237 DLTSASYTMI 246
 DB 268 DLTSASYTMI 277

RESULT 29

ID ABR56317 standard; protein; 250 AA.
 AC ABR56317;
 DT 20-NOV-2003 (first entry)

DE Human galectin-3.
 DE Human; cytosolic; galectin-3; cancer.

OS Homo sapiens.

Key Location/Qualifiers
 FT 117..250
 FT Domain /label=Carbohydrate_recognition_domain

PN US2003054982-A1.

PD 20-MAR-2003.

PF 08-JUN-2001; 2001US-00877790.

PR 08-JUN-2001; 2001US-00877790.

PA (JARV/) JARVIS G A.

PA (JOHN/) JOHN C M.

PA (LEFF/) LEFFLER H.

PI Jarvis GA, John CM, Leffler H;

DR WPI; 2003-439728/41.

XX New composition, useful for treating cancer, comprises N-terminally

PT truncated galectin-3 and a carrier.

XX Disclosure; Page 4; 30pp; English.

CC The present invention relates to a composition (I), comprising an amount

CC of N-terminally truncated galectin-3 (ABR56316) and a carrier. The

CC composition is used for treating cancer. The present sequence is the full

CC -length human galectin-3 sequence

XX Sequence 250 AA;

Query Match 94.8%; Score 1287; DB 6; Length 250;

Best Local Similarity 96.8%; Pred. No. 2.9e-94;

Matches 242; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

DB 1 MADNFSLHDALSSGNGNPNPGWPGAMGNOGAPAGAGYFGASYPGAYPGQAPP 60
 QY 60 GAYHGAAPGAPGAPAPGVYPPGPPGAPYSSSGSPAPGAY-ATGPGYGAAPGLIVRYNL 118
 DB 61 GAYHGAAPGAPGAPAPGVYPPGPPGAPYSSSGSPAPGAYPATGPGYGAAPGLIVRYNL 120
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHF-PRFNNRNRYIVCNTKLDNN 177
 DB 121 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHFNNRNRYIVCNTKLDNN 190
 QY 178 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAH-LQYNHRVKYKLINEISKIGISGDI 236
 DB 181 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAHLLQYNHRVKYKLINEISKIGISGDI 240
 QY 237 DLTSASYTMI 246
 DB 241 DLTSASYTMI 250

RESULT 30

ID AAR12532 standard; protein; 248 AA.
 AC AAR12532;
 DT 25-MAR-2003 (revised)

DT 12-SEP-1991 (first entry)

DE Human Macrophage Carbohydrate- and Ige-binding protein.

DE HMEBP; leishmaniasis; Mouse Mac-2; laminin.

OS Homo sapiens.

Key Location/Qualifiers

FT 50..55
 FT Region /note="repeated 5 times"

FT 59..64
 FT Region /note="repeated 5 times"

FT 65..71
 FT Region /note="repeated 5 times"

FT 75..80
 FT Region /note="repeated 5 times"

FT 96..101
 FT Region /note="repeated 5 times"

FT 156..161
 FT Region /note="feature indicated but not described"

FT 179..184
 FT Region /note="feature indicated but not described"

PN MO9108290-A.

PD 13-JUN-1991.

PF 30-NOV-1989; 89US-00444195.

PR 30-NOV-1989; 89US-00444195.

PR 14-SEP-1990; 90US-00582628.

PA (GEHO) GEN HOSPITAL CORP.

PI Pillal S, Cherayil BJ;

DR WPI; 1991-1919196/26.

DR N-PSDB; AAQ12211.

XX Recombinant gene encoding human macrophage carbohydrate - Ige-binding

XX protein and antibody used to treat, diagnose and prevent e.g.

XX inflammatory bowel disorder, leishmaniasis, hay fever, etc.

PS Claim 1; Fig 3; 36pp; English.

XX This sequence is deduced from a composite of two overlapping cDNA clones

CC isolated from a human colon cancer cell line HT-29 cDNA library. The
CC deduced carboxy-terminal half of HMEBP resembles that deduced for mouse
CC Mac-2. There is more divergence in the amino-terminal portion. The
CC estimated size of this protein is 2x0 smaller than that of its mouse
CC counterpart. The two proteins exhibit 77 per cent amino acid sequence
CC identity. See also AA012207-012210. (Updated on 25-MAR-2003 to correct PA
CC field.)
XX

SQ Sequence 248 AA;

Query Match 88.0%; Score 1194; DB 2; Length 248;
Best Local Similarity 92.4%; Pred. No. 6.9e-87;
Matches 230; Conservative 0; Mismatches 15; Indels 4; Gaps 4;

QY 1 MADNFSLHDLSSGSGNPQGMWPGAMGNDPAGAGYPPGAYPGQAPPG 60
DB 1 MADNFSLHDLSSGSGNPQGMWPGAMGNDPAGAGYPPGAYPGQAPPG 60
QY 61 AYHGAAPGAPGAPGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119
DB 61 AYHGAAPGAPGAPGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119
QY 120 LFGGVPRMLITLIGTVKPNANRIALDFQGNDAVFHF-PRENNRRVIVCNTKLDNNW 178
DB 120 LFGGVPRMLITLIGTVKPNANRIALDFQGNDAVFHF-PRENNRRVIVCNTKLDNNW 178
QY 179 GREEROSVFPFESGKPKIQVIVEDHFKVAVNDAH-LQYHNVKKNLEISKLGISGID 237
DB 180 GREEROSVFPFESGKPKIQVIVEDHFKVAVNDAH-LQYHNVKKNLEISKLGISGID 237
QY 238 LTSASYTMI 246
DB 240 LTSASYTMI 248

RESULT 31

ABU04682
ID ABU04682 standard; protein; 248 AA.

AC ABU04682;

DT 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #1348.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

XX WO200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCOS INC.

PI Chicx RM, Tomlinson AJ, Urban RG;

DR WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.

PS Example 2; SEQ ID NO 1348; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease or protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 248 AA;

Query Match 88.0%; Score 1194; DB 6; Length 248;
Best Local Similarity 92.4%; Pred. No. 6.9e-87;
Matches 230; Conservative 0; Mismatches 15; Indels 4; Gaps 4;

QY 1 MADNFSLHDLSSGSGNPQGMWPGAMGNDPAGAGYPPGAYPGQAPPG 60
DB 1 MADNFSLHDLSSGSGNPQGMWPGAMGNDPAGAGYPPGAYPGQAPPG 60
QY 61 AYHGAAPGAPGAPGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119
DB 61 AYHGAAPGAPGAPGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119
QY 120 LFGGVPRMLITLIGTVKPNANRIALDFQGNDAVFHF-PRENNRRVIVCNTKLDNNW 178
DB 120 LFGGVPRMLITLIGTVKPNANRIALDFQGNDAVFHF-PRENNRRVIVCNTKLDNNW 178
QY 179 GREEROSVFPFESGKPKIQVIVEDHFKVAVNDAH-LQYHNVKKNLEISKLGISGID 237
DB 180 GREEROSVFPFESGKPKIQVIVEDHFKVAVNDAH-LQYHNVKKNLEISKLGISGID 237
QY 238 LTSASYTMI 246
DB 240 LTSASYTMI 248

RESULT 32

AAR12531
ID AAR12531 standard; protein; 264 AA.

AC AAR12531;

DT 25-MAR-2003 (revised)

DT 12-SEP-1991 (first entry)

DE Mac2.16 expression product.

XX HMEBP; leishmaniasis; Mouse Mac-2; laminin.

XX Mus musculus.

XX WO9108290-A.

XX 13-JUN-1991.

PF 30-NOV-1989; 89US-00444195.

PR 30-NOV-1989; 89US-00444195.

[illegible]

XX (PROT-) PROTEGENE KK.
XX
XX WPI; 2003-367092/35.
DR
DR N-PSDB; ADCS3850.
XX
XX Agents for prevention and treatment of nephritis, comprise galectin-1,
PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration
PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8
PT positive cells.
XX
XX Claim 1; SEQ ID NO 2; 31pp; Japanese.
PS
XX This invention relates to the use of novel mammal derived galectin-1
XX (G3), -3 (G3) and -9 (G9) proteins as effective ingredients for
CC (G3), -3 (G3) and -9 (G9) proteins as effective ingredients for
CC prevention and treatment of nephritis. The invention discloses agents for
CC prevention and treatment of nephritis, particularly glomerular nephritis
CC prevention and treatment of nephritis, particularly glomerular nephritis
CC and may have antiinflammatory activities. The method of the invention
CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and
CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and
CC apoptosis of CD8 positive cells. The method and sequences of the
CC invention may be used for prevention and treatment of nephritis,
CC particularly glomerular nephritis including inhibition of glomerular
CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8
CC positive cells. The present sequence represents the mouse galectin 3
CC protein of the invention.
XX
XX Sequence 264 AA;
SQ

[illegible]

Accession	Protein	Location/Qualifiers
AA133338	AA133338 standard; protein; 278 AA.	
AA133338	AA133338;	
XX		
AC		
XX	25-MAR-2003 (revised)	
DT	12-SEP-1991 (first entry)	
DT		
XX	Mac-2 protein including putative signal sequence.	
DE		
XX		
KW	HMEBP; leishmaniasis; Mouse Mac-2; laminin.	
XX		
OS	Mus musculus.	
XX		
FH	Key	
FT	Peptide	
FT		1..14
FT	/label= putative signal peptide	
FT	/note= "Met(1) is encoded by CTG	
FT	15..278	
FT	/label= Mac-2	

XX
PN WO9108290-A.
XX
PD 13-JUN-1991.
XX
PF 30-NOV-1989; 89US-00444195.
XX
PR 30-NOV-1989; 89US-00444195.
XX
PR 14-SEP-1990; 90US-00582628.
XX
PA (GEO) GEN HOSPITAL CORP.
XX
PI Pillai S, Cherayil BJ;
XX
DR WPI: 1991-193196/26.
XX
DR N-PSDB; AAQ12210.
XX
PT Recombinant gene encoding human macrophage carbohydrate - IgE-binding
XX
PT protein and antibody used to treat, diagnose and prevent e.g.
XX
PT inflammatory bowel disorder, leishmaniasis, hay fever, etc.
XX
PS Disclosure; Fig 2D, 36pp; English.
XX
CC This sequence is identical to that deduced from clone Mac 2.16 (see
XX
CC AAR12531), except for the putative signal peptide. Mac 2.16 itself does
XX
CC not encode a signal peptide, but Mac-2 protein newly synthesised by mouse
XX
CC inflammatory peritoneal macrophages was detected in the extracellular
XX
CC medium. Clone Mac 2.9 (from which this sequence was deduced) was thus
XX
CC analysed for presence of a signal sequence. See also AAQ12207 (for Mac
XX
CC 2.16), AAQ12209, AAQ12211. (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 278 AA;

Query Match 79.7%; Score 1082; DB 2; Length 278;
Best Local Similarity 78.0%; Pred. No. 6.1e-78;
Matches 209; Conservative 16; Mismatches 17; Indels 26; Gaps 8;

QY 1 MADNFSLHDLAAGSGNPNPQGWPGAWGQAPGAGGYPGASYPG-YPGQAPPGAYPGQAP 59
DB 15 MADNFSLHDLAAGSGNPNPQGWPGAWGQAP-GAGGYPGASYPGAYPGQAPPGAYPGQAP 73
QY 60 GAYHG-----APGAYPGAPAGVYPPGSPGAYSSGQPSAPGAY----- 100
DB 74 GAYGQAPPGAYPGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130
QY 101 ATGPGAPAGPLIVPNPLPFGVVPRLITLITGVKNANRILDPORGNDVAFHF-PR 159
DB 131 AAGGYGVPAGPLIVPNPLPFGVVPRLITLITGVKNANRILDPORGNDVAFHF-PR 190
QY 160 FENNNRRVIVCNTKLDNNWGEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAAH-LQYN 218
DB 191 FENNNRRVIVCNTKLDNNWGEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAAH-LQYN 250
QY 219 HRVKKLNEISKLGISGIDILTSASYTM 246
DB 251 HRMKNLREISQLGISGIDILTSASHAMI 278

RESULT 35
AAU97821
ID AAU97821 standard; protein; 262 AA.
AC AAU97821;
XX
XX
DT 27-AUG-2002 (first entry)
XX
DE Rat cell membrane anchor protein galectin-3.
XX
XX
KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;
KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
KW polycystic kidney; post-angioplasty restenosis; cytostatic;

KW Immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
KW vasotropic; hepatotropic; rat.
XX
OS Rattus sp.
XX
PN WO200229031-A2.
XX
PD 11-APR-2002.
XX
PF 01-OCT-2001; 2001WO-IL000918.
XX
PR 04-OCT-2000; 2000US-0237858P.
XX
PR (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PA KLoog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
XX
PI WPI: 2002-43533/46.
XX
DR Identifying anchor proteins that bind Ras protein, by producing complexes
XX
DR of Ras and cell membrane proteins in the presence and absence of a Ras
XX
PT antagonist and identifying a complex disrupted by the Ras antagonist.
XX
PS Disclosure; Page 14; 62pp; English.
XX
XX
CC The invention describes a method of identifying cell membrane anchor
XX
CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
XX
CC where one mixture has a Ras antagonist. A cross linking agent is added,
XX
CC and complexes between Ras protein and other proteins are produced. The
XX
CC complexes are then separated and the proteins binding to Ras are
XX
CC identified. The invention also describes a method useful for identifying
XX
CC drug candidates that inhibit aberrant Ras activity. An antisense compound
XX
CC comprising at least one phosphorothioate-modified nucleotide is useful
XX
CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
XX
CC compound into a patient exhibiting this problem. The method is also
XX
CC useful for identifying anchor proteins for the farnesylated isoforms of H
XX
CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
XX
CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
XX
CC for treating diseases characterised by uncontrolled mitosis, including
XX
CC cancers and various non-malignancies such as autoimmune disease (e.g.
XX
CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
XX
CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty
XX
CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
XX
XX
SQ protein that binds an isoform of Ras

Query Match 79.5%; Score 1079; DB 5; Length 262;
Best Local Similarity 77.9%; Pred. No. 9.9e-78;
Matches 205; Conservative 16; Mismatches 24; Indels 18; Gaps 7;

QY 1 MADNFSLHDLAAGSGNPNPQGWPGAWGQAPGAGGYPGASYPG-YPGQAPPGAYPGQAP 59
DB 1 MADNFSLHDLAAGSGNPNPQGWPGAWGQAP-GAGGYPGASYPGAYPGQAPPGAYPGQAP 59
QY 60 GAYHG-----APGAYPGAPAGVYPPGSPGAYPSS--GQSPAPGAY-ATGPY 105
DB 60 SATPDPGPGSAYPGPTAPGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 119
QY 106 GATPGLIVPNPLPFGVVPRLITLITGVKNANRILDPORGNDVAFHF-PR-RENN 164
DB 120 GATPGLIVPNPLPFGVVPRLITLITGVKNANRILDPORGNDVAFHF-PR-RENN 179
QY 165 RRVIVCNTKLDNNWGEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAAH-LQYNRVK 223
DB 180 RRVIVCNTKLDNNWGEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDVHLLQYNRMKN 239
QY 224 LNEISKLGISGIDILTSASYTM 246
DB 240 LNEISKLGISGIDILTSASHAMI 262

RESULT 36

ADCS3847
 ID ADCS3847 standard; protein, 262 AA.
 XX
 AC ADCS3847;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Rat galectin 3 protein.
 XX
 DE galectin-3; nephritis; glomerular nephritis; antiinflammatory;
 KW glomerular infiltration; apoptosis; Rat.
 XX
 OS Rattus norvegicus.
 XX
 FN JP2002322082-A.
 XX
 PD 08-NOV-2002.
 XX
 PF 26-APR-2001; 2001JP-00129200.
 XX
 PR 26-APR-2001; 2001JP-00129200.
 XX
 PA (PROT-) PROTEGENE KK.
 XX
 PA WPI; 2003-367092/35.
 DR
 XX
 XX Agents for prevention and treatment of nephritis, comprise galectin-1,
 PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration
 PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8
 PT positive cells.
 XX
 PS Disclosure; SEQ ID NO 8; 31pp; Japanese.
 XX
 CC This invention relates to the use of novel mammal derived galectin-1
 CC (G1), -3 (G3) and -9 (G9) proteins as effective ingredients for
 CC prevention and treatment of nephritis. The invention discloses agents for
 CC prevention and treatment of nephritis, particularly glomerular nephritis
 CC and may have antiinflammatory activities. The method of the invention
 CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and
 CC apoptosis of CD8 positive cells. The method and sequences of the
 CC invention may be used for prevention and treatment of nephritis,
 CC particularly glomerular nephritis including inhibition of glomerular
 CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8
 CC positive cells. The present sequence represents the Rat galectin 3
 CC protein used in the exemplification of the method of the invention.
 CC
 SQ Sequence 262 AA;

Query Match 79.5%; Score 1079; DB 7; Length 262;
 Best Local Similarity 77.9%; Pred. No. 9.9e-78; Indels 18; Gaps 7;
 Matches 205; Conservative 16; Mismatches 24;

QY 1 MADNFSIHDALSSGNNPQGWPGAMGNCPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
 DB 1 MADNFSIHDALSSGNNPQGWPGAMGNCPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
 QY 60 GAYHG-----APGAYPGAPAPGAYPPGPGSGGAYPSS--GQPSAPGAY-ATGPGY 105
 DB 60 SAYPGPTGPGAPYDPTAPGAYPGPTAPGAPPGQPGGAYPSAPGAYPSAPGAYPATGPGF 119
 QY 106 GARGPLIVRYNPLPGGVPRMLITLIGTVKPNANRIALDFORGNVAFHF-PRFENN 164
 DB 120 GARGPLIVRYNPLPGGVPRMLITLIGTVKPNANRIALDFORGNVAFHF-PRFENN 179
 QY 165 RRVIVCTKLDNNNGREEROSVFPFESGKPFKIQVLVEPDHFKVAVNDAA-LQYNHRVK 223
 DB 180 RRVIVCTKLDNNNGREEROSVFPFESGKPFKIQVLVEADHFKVAVNDVHLQYNHRMKN 239
 QY 224 LNEISKLGISGIDILTSASYTMI 246
 DB 240 LREISQLGISGIDILTSASHAMI 262

RESULT 37
 AAW71219
 ID AAW71219 standard; protein, 264 AA.
 XX
 AC AAW71219;
 XX
 DT 30-OCT-1998 (first entry)
 XX
 DE Beta-D-galactoside-binding protein designated L-31-gal-lectin.
 XX
 DE Beta-D-galactoside-binding protein; L-31-gal-lectin;
 KW metastatic potential; antibody.
 XX
 OS Homo sapiens.
 XX
 FN US5801002-A.
 XX
 PD 01-SEP-1998.
 XX
 PF 22-NOV-1995; 95US-00562311.
 XX
 PR 06-JAN-1989; 89US-00294249.
 PR 05-APR-1991; 91US-00681242.
 PR 26-JAN-1994; 94US-00188225.
 XX
 PA (KARM-) KARMANOS CANCER INST BARBARA ANN.
 XX
 PA Raz A;
 XX
 DR WPI; 1998-494766/42.
 DR N-PSDB; AAV54736.
 XX
 XX Test for metastatic potential of cell sample - by measuring binding of
 PT antibody to L-31-gal-lectin on cell surface.
 PT
 PS Disclosure; Fig 3A-B; 24pp; English.
 XX
 CC The present sequence represents a beta-D-galactoside-binding protein
 CC designated L-31-gal-lectin. The sequence is derived from clone 2. The
 CC specification describes a method for testing a cell sample for metastatic
 CC potential. The method comprises contacting the sample with a labelled
 CC antibody that binds to endogenous cell-surface L-31-gal-lectin, removing
 CC unbound antibody, and determining the amount of bound antibody as a
 CC measure of L-31-gal-lectin expression, where the metastatic potential
 CC increases as the level of L-31-gal-lectin expression increases. The
 CC antibody is produced by immunisation with a L-31-gal-lectin protein
 CC
 SQ Sequence 264 AA;

Query Match 79.4%; Score 1078; DB 2; Length 264;
 Best Local Similarity 78.2%; Pred. No. 1.2e-77; Indels 22; Gaps 8;
 Matches 208; Conservative 15; Mismatches 21;

QY 1 MADNFSIHDALSSGNNPQGWPGAMGNCPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
 DB 1 MADNFSIHDALSSGNNPQGWPGAMGNCPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
 QY 60 GAYHG--APGAYPGAPAPGAYPPGPGSGGAYPSS-----GQPSAPGAY-----AT 102
 DB 60 GAYPGAPAPGAYPPGPGSGGAYPSS-----GQPSAPGAY-----AT 118
 QY 103 GPGGAPGGLIVRYNPLPGGVPRMLITLIGTVKPNANRIALDFORGNVAFHF-PRFN 161
 DB 119 GPGGAPGGLIVRYNPLPGGVPRMLITLIGTVKPNANRIALDFORGNVAFHF-PRFN 178
 QY 162 ENNRVRVCTKLDNNNGREEROSVFPFESGKPFKIQVLVEPDHFKVAVNDAA-LQYNHR 220
 DB 179 ENNRVRVCTKLDNNNGREEROSVFPFESGKPFKIQVLVEADHFKVAVNDVHLQYNHR 238
 QY 221 VKKLEISKLGISGIDILTSASYTMI 246
 DB 239 MKKLEISQLGISGIDILTSANHAMI 264

```
RESULT 38
ABU04782
ID ABU04782 standard; protein; 264 AA.
XX
AC ABU04782;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1448.
XX
KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200278524-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WC-US009671.
XX
PR 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985E.
XX
PA (ZYCO-) ZYCOs INC.
XX
PI Chicx RM, Tomlinson AJ, Urban RG;
XX
DR WPI; 2003-040607/03.
XX
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX
PS Example 2; SEQ ID NO 1448; 134pp; English.
XX
XX
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. There are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 264 AA;
Query Match 79.4%; Score 1078; DB 6; Length 264;
Best Local Similarity 78.2%; Pred. No. 1.2e-77;
Matches 208; Conservative 15; Mismatches 21; Indels 22; Gaps 8;
```

```
QY 60 GAYHG--APGAYGAPAGVPPGPPSGAGAYSS-----GQSAAGAY-----AT 102
DB 60 GAYRGAPPSAIPGCPAPAPAPG-PTAPAPVPGSTAPGAFPGQPGAPGAYPSAPGAYPAA 118
QY 103 GPYGAPAGPLIPEYMLPLPGVPPRLITLITGVKPNARIALDPGRGNDVAFHF-PRFN 161
DB 119 GPYGVAPGLTVPYDPLPGMLPRLLITMTGVKNARIYVDFRGNDVAFHFPRFN 178
QY 162 ENNRRTIVNTKLDNNWGKEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAA-LOYNHR 220
DB 179 ENNRRTIVNTQDDNNWGKEERQSAFPFSGKPFKIQVLVEADHFKVAVNDALLQYNHR 238
QY 221 VKLNRISKLGISGIDILTSASYTMI 246
DB 239 MGLNRISQLGISGIDILTSANHAMI 264
```

```
RESULT 39
AAR42200
ID AAR42200 standard; protein; 262 AA.
XX
AC AAR42200;
XX
DT 25-MAR-2003 (revised)
DT 18-MAY-1994 (first entry)
XX
DE IGE binding protein.
XX
KW IGE; immunoglobulin; binding protein; allergy; allergic reaction;
KW absorbance; regulation; immune response.
XX
OS Rattus rattus.
XX
PN US5260434-A.
XX
PD 09-NOV-1993.
XX
PF 10-JUL-1991; 91US-00728125.
PR 23-APR-1985; 85US-00726250.
PR 21-SEP-1988; 88US-00247170.
XX
PA (SCRI ) SCRIPPS RES INST.
XX
PI Liu F;
XX
DR WPI; 1993-367951/46.
DR N-PSDB; AA050878.
XX
PT DNA encoding IGE-binding protein - with repetitive sequence and homology
PT with IGE receptor.
XX
PS Claim 1; Fig 1; 20pp; English.
XX
CC The Ige binding protein is used to regulate or absorb IGE. It is useful
CC particularly in the absorbance of IGE to ease allergic reactions.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 262 AA;
Query Match 79.2%; Score 1075; DB 2; Length 262;
Best Local Similarity 77.6%; Pred. No. 2e-77;
Matches 204; Conservative 17; Mismatches 24; Indels 18; Gaps 7;
```

Db 120 GAPTEPLTPVPMPLPGVMPRLITITIGVKNPNSITLTFKKNDAFHENPRENN 179
 QY 165 RRVICNTKLDNNWGEROSVFPESGKPFKIQVLVEPDHFKVAVNDH-LQYNHRVK 223
 Db 180 RRVICNTKODNNWGEROSAFPPESGKPFKIQVLVEADHFKVAVNDVHLQYNHRKKN 239
 QY 224 INEISKLGISGDIDITLSASYMI 246
 Db 240 REISQLGIIIGDITLTSASHAMI 262
 RESULT 40
 ADE56586
 ID ADE56586 standard; protein; 261 AA.
 AC ADE56586;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein P08699, SEQ ID NO 2440.
 XX
 KM Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KM chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PI Woolf C, D'urso D, Befort K, Costigan M;
 PT WPI: 2003-268312/26.
 PT GENBANK; P08699.
 PS New composition comprising two or more isolated polypeptides, useful for
 preparing a medicament for treating pain in an animal.
 PS Claim 1; Page; 1017pp; English.

CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp://ipo.int/pub/published_pct_sequences.
 SQ Sequence 261 AA;
 Query Match 79.1%; Score 1074; DB 7; Length 261;
 Best Local Similarity 77.9%; Pred. No. 2,4e-77;
 Matches 204; Conservative 16; Mismatches 24; Indels 18; Gaps 7;
 QY 2 ADNPSLHDALSGSGNPNQMPGANGNPAGAGTGGASYRG-YGGAPPAGYPPQAPPG 60
 Db 1 ADGFSLNALAGSGNPNQMPGAMGNP-GAGTGGASYRGAYPGAPPGYPPQAPPS 59
 QY 61 AYHG-----APGAYCAPAPGAYPPGPPSGRGAYPPS--GQPSAPGAY-ATGPGY 106
 Db 60 AYPGPTGSAVYGPAPAPAYGPTAPGAPPGQPGGAPYAPGAPYAPGAPYATGPG 119
 QY 107 APAGPLIYPVNLPLPGVMPRLITITIGVKNPNSITLTFKKNDAFHENPRENNR 165
 Db 120 APTGPLTPVPMPLPGVMPRLITITIGVKNPNSITLTFKKNDAFHENPRENNR 179
 QY 166 RRVICNTKLDNNWGEROSVFPESGKPFKIQVLVEPDHFKVAVNDH-LQYNHRVK 224
 Db 180 RRVICNTKODNNWGEROSAFPPESGKPFKIQVLVEADHFKVAVNDVHLQYNHRKKN 239
 QY 225 NEISKLGISGDIDITLSASYMI 246
 Db 240 REISQLGIIIGDITLTSASHAMI 261
 RESULT 41
 AAU97822
 ID AAU97822 standard; protein; 263 AA.
 AC AAU97822;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE Mouse cell membrane anchor protein galectin-3.
 XX
 KM Galectin-3; cell membrane anchor protein; Ras; antisense technology;
 KM farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
 KM cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
 KM multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
 KM polycystic kidney; post-angioplasty restenosis; cytostatic;
 KM immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
 KM vasotrophic; hepatotropic; mouse.
 XX
 OS Mus sp.
 XX
 PN WO200229031-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 01-OCT-2001; 2001WO-IL000918.
 PR 04-OCT-2000; 2000US-0237858P.
 PA (UYRA-) UNIV RAVOT APPLIED RAS & IND DEV LTD.
 PI KIoog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
 PT WPI: 2002-435333/46.
 PT Identifying anchor proteins that bind Ras protein, by producing complexes
 of Ras and cell membrane proteins in the presence and absence of a Ras
 antagonist and identifying a complex disrupted by the Ras antagonist.
 PS Disclosure; Page 14; 62pp; English.
 CC The invention describes a method of identifying cell membrane anchor

XX WO2002100343-A2.
 XX 19-DEC-2002.
 XX 10-JUN-2002; 2002WO-US018478.
 XX 08-JUN-2001; 2001US-0296970P.
 XX (MAND-) MANDALMED INC.
 XX JARVIS GA, John CM, Leffler H;
 XX WPI, 2003-183915/18.
 XX Novel N-terminally truncated galectin-3 derivatized with polyethylene
 XX glycol, useful for treating tumor.
 XX Claim 2; Page 22; 88pp; English.
 XX The sequences given in AAG79833-35 represent recombinant human galectin-
 XX 3's. Galectins bind lactose and the human protein is composed of 250
 XX amino acids with a molecular weight of approx. 31000. N-terminally
 XX truncated galectin-3 derivatized with at least one molecule of
 XX polyethylene glycol is useful for treating a tumor in a patient and for
 XX reducing tumour size, and for treating cancer or for preventing
 XX metastasis. It is also useful for inhibiting carbohydrate binding and
 XX multimerisation of galectin-3, and for inhibiting tumour growth and
 XX metastasis in vivo. DNA encoding the truncated galectin-3 is useful in
 XX gene therapy
 XX Sequence 143 AA;

Query Match 53.3%; Score 723; DB 6; Length 143;
 Best Local Similarity 98.6%; Pred. No. 9.3e-50;
 Matches 141; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 106 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFORGNDVAFHF-PRNENN 164
 DB 1 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFORGNDVAFHFPRNENN 60
 QY 165 RRVYCNCTLDNNMGRERQSVFPESGKPKIQVLEBDHFKVAVNDAH-LQYNHRVK 223
 DB 61 RRVYCNCTLDNNMGRERQSVFPESGKPKIQVLEBDHFKVAVNDAHLLQYNHRVK 120
 QY 224 LNEISKLGISGIDILTSASYTMI 246
 DB 121 LNEISKLGISGIDILTSASYTMI 143

RESULT 44
 AAG79834
 ID AAG79834 standard; protein; 144 AA.
 XX AAG79834;
 XX 16-APR-2003 (first entry)
 XX N-terminally truncated Galectin-3 with additional N-terminal Cys.
 XX Galectin-3; nuclear localisation; Bcl-2; metastasis;
 XX N-terminally truncated; polyethylene glycol; tumour; cancer;
 XX carbohydrate binding; multimerisation; gene therapy.
 XX Homo sapiens.
 XX WO2002100343-A2.
 XX 19-DEC-2002.
 XX 10-JUN-2002; 2002WO-US018478.
 XX 08-JUN-2001; 2001US-0296970P.

XX (MAND-) MANDALMED INC.
 XX JARVIS GA, John CM, Leffler H;
 XX WPI, 2003-183915/18.
 XX Novel N-terminally truncated galectin-3 derivatized with polyethylene
 XX glycol, useful for treating tumor.
 XX Disclosure; Page 22; 88pp; English.
 XX The sequences given in AAG79833-35 represent recombinant human galectin-
 XX 3's. Galectins bind lactose and the human protein is composed of 250
 XX amino acids with a molecular weight of approx. 31000. N-terminally
 XX truncated galectin-3 derivatized with at least one molecule of
 XX polyethylene glycol is useful for treating a tumor in a patient and for
 XX reducing tumour size, and for treating cancer or for preventing
 XX metastasis. It is also useful for inhibiting carbohydrate binding and
 XX multimerisation of galectin-3, and for inhibiting tumour growth and
 XX metastasis in vivo. DNA encoding the truncated galectin-3 is useful in
 XX gene therapy
 XX Sequence 144 AA;

Query Match 53.3%; Score 723; DB 6; Length 144;
 Best Local Similarity 98.6%; Pred. No. 9.4e-50;
 Matches 141; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 106 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFORGNDVAFHF-PRNENN 164
 DB 2 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFORGNDVAFHFPRNENN 61
 QY 165 RRVYCNCTLDNNMGRERQSVFPESGKPKIQVLEBDHFKVAVNDAH-LQYNHRVK 223
 DB 62 RRVYCNCTLDNNMGRERQSVFPESGKPKIQVLEBDHFKVAVNDAHLLQYNHRVK 121
 QY 224 LNEISKLGISGIDILTSASYTMI 246
 DB 122 LNEISKLGISGIDILTSASYTMI 144

RESULT 45
 ABR56316
 ID ABR56316 standard; protein; 141 AA.
 XX ABR56316;
 XX 20-NOV-2003 (first entry)
 XX N-terminally truncated human galectin-3.
 XX Human; cytosstatic; galectin-3; cancer.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 XX Domain 10..141
 XX /label= Carbohydrate_recognition_domain
 XX US2003054982-A1.
 XX 20-MAR-2003.
 XX 08-JUN-2001; 2001US-00877790.
 XX 08-JUN-2001; 2001US-00877790.
 XX (JARV/) JARVIS G A.
 XX (JOHN/) JOHN C M.
 XX (LEFF/) LEFFLER H.
 XX JARVIS GA, John CM, Leffler H;

XX WPI; 2003-439728/41.
XX
XX
PT New composition, useful for treating cancer, comprises N-terminally
truncated galectin-3 and a carrier.
XX
XX
PS Claim 2; Page 4; 30pp; English.
XX
CC The present invention relates to a composition (I), comprising an amount
of N-terminally truncated galectin-3 (ABR56316) and a carrier. The
composition is used for treating cancer. The present sequence was
produced by exhaustive digestion with collagenase
CC
XX Sequence 141 AA;
SQ
Query Match 52.3%; Score 710; DB 6; Length 141;
Best Local Similarity 97.2%; Pred. No. 9.9e-49;
Matches 138; Conservative 1; Mismatches 1; Indels 2; Gaps 2;
QY 106 GAPAGPLIVPYNLPLPGGVPRMLITITIGVKNPNANRIALDFORGNDVAFHF-PRFNNENN 164
DB 1 GAPAGPLIVPYNLPLPGGVPRMLITITIGVKNPNANRIALDFORGNDVAFHFNNRNNENN 59
QY 165 RRVIVCNKLDNNMGRERQSVFPESGKPEKIQVLVEPDHFKVAVNDALH-QYNHRVKKL 224
DB 60 RRVIVCNKLDNNMGRERQSVFPESGKPEKIQVLVEPDHFKVAVNDALH-QYNHRVKKL 119
QY 225 NEISKLGISGIDILTSASYTMI 246
DB 120 NEISKLGISGIDILTSASYTMI 141
RESULT 46
AAP60534
ID AAP60534 standard; protein; 138 AA.
XX
AC AAP60534;
XX
DT 03-OCT-2002 (revised)
DT 01-JAN-1980 (first entry)
XX
DE C-terminal of IGE-binding protein.
XX
KM IGE; immunoglobulin; binding protein; antiallergic; ss.
XX
OS Rattus rattus.
XX
PN WO8606407-A.
XX
PD 06-NOV-1986.
XX
PF 18-APR-1986; 86WO-US000810.
XX
PR 23-APR-1985; 85US-00726250.
XX
PA (MEDI-) MEDICAL BIOLOGY INS.
XX
PI Lu FT;
XX
DR WPI; 1986-305131/46.
DR N-PSDB; AAN60456.
XX
PT New IGE-binding polypeptide and its CDNA - may be useful in handling of
allergic reactions.
XX
PS Disclosure; Fig 3; 32pp; English.
XX
CC The IGE-binding protein partially encoded by this C-terminal sequence may
be used to establish the structural relatedness of the IGE systems
proteins. Specifically, it may be used as an IGE regulant or to absorb
IGE in the handling of allergic reactions. (Updated on 03-OCT-2002 to
add missing OS field.)
CC
XX

SQ Sequence 138 AA;
Query Match 43.7%; Score 593; DB 1; Length 138;
Best Local Similarity 81.9%; Pred. No. 1.9e-39;
Matches 113; Conservative 11; Mismatches 12; Indels 2; Gaps 2;
QY 111 PLIVPYNLPLPGGVPRMLITITIGVKNPNANRIALDFORGNDVAFHF-PRFNNNRVIV 169
DB 1 PLIVPYNLPLPGGVPRMLITITIGVKNPNANRIALDFORGNDVAFHFNNRNNRVIV 60
QY 170 CNTKLDNNMGRERQSVFPESGKPEKIQVLVEPDHFKVAVNDALH-QYNHRVKKLDEL 228
DB 61 CNTKLDNNMGRERQSVFPESGKPEKIQVLVEPDHFKVAVNDALH-QYNHRVKKLDEL 120
QY 229 KLGISGIDILTSASYTMI 246
DB 121 QLGITIGDILTSASYTMI 138
RESULT 47
AAB97171
ID AAB97171 standard; protein; 341 AA.
XX
AC AAB97171;
XX
DT 13-AUG-2001 (first entry)
XX
DE Rainbow trout galectin.
XX
KM Rainbow trout; galectin; cytostatic; antiinflammatory; immunity;
KW fish disease; cell adhesion inhibition; inflammation; cancer;
XX T cell movement; apoptosis.
XX
OS Oncorhynchus mykiss.
XX
PN JP2001069976-A.
XX
PD 21-MAR-2001.
XX
PF 01-SEP-1999; 99JP-00247204.
XX
PR 01-SEP-1999; 99JP-00247204.
XX
PA (NORQ) NORINSUISANSO YOSHOKU KENKYU.
XX
DR WPI; 2001-321173/34.
DR N-PSDB; AAH24641.
XX
PT Novel recombinant Rainbow trout galectin protein and gene encoding the
protein, useful for studying fish immunity mechanism, and diagnosis of
fish diseases.
XX
PS Claim 1; Page 10-11; 17pp; Japanese.
XX
CC The invention relates to a recombinant protein having a 341 amino acid
sequence fully defined in the specification, or its mutant in which at
least one amino acid is deleted, substituted or added, but which retains
the galectin activity. The protein is useful for studying mechanisms of
fish immunity and for diagnosing fish diseases. The galectin proteins are
involved in inhibition of cell adhesion, inflammation, metastasis of
tumour cells, T cell movement, and apoptosis. The present sequence is the
protein of the invention
CC
XX
SQ Sequence 341 AA;
Query Match 23.2%; Score 314.5; DB 4; Length 341;
Best Local Similarity 36.0%; Pred. No. 6.4e-17;
Matches 90; Conservative 30; Mismatches 69; Indels 61; Gaps 17;
QY 9 DALSGSG-----NPNP-----QGWGAMGNQAGAGYFGASYPGQAPPGAY 53
DB 132 DTISADGKVELTIVFONPAPPTIPADGPFA-----QP-----GFP--SYPGFPAQ--PG-F 178

```

QY 54 EQGAPGAYHGAAPGAPGAYGPPSPGSGAPSSGQSPAPGAYATGPY-GAPAGPL 112
DB 129 P-----SYPGFPA---QPGFSPG-PP--GQPGF-----PYPGFPAQ- 211
QY 113 IYPYNIPLPGVGVPMILITILGTVPKPNARIALDFORGNDVAFHF-PRENENRRVIYCN 171
DB 212 AVFYNNMNGIYFPORTINIQGVNPNANRFFHNLFFNSGIALHFNPRDEF---LVYEN 268
QY 172 TKLDNNKREERQSYVFPFSGKPEFKIQVLVEPDHFKAVDNHLQ-YNHRVKKLINEISKI 230
DB 269 SKLRDQWKEERSGGWPFHFGQAFILITICDAQCYKIVNGNQTSTYKRRHTLLQOVNLL 328
QY 231 GTSDDILTS 240
DB 329 EVDGDLSTS 338

RESULT 48
ADC53842
ID ADC53842 standard; protein; 322 AA.
XX ADC53842;
AC
XX
XX 18-DEC-2003 (first entry)
DT
XX
DE Mouse galectin 9 protein.
KW galectin-9; nephritis; glomerular nephritis; antinflammatory;
KW glomerular infiltration; apoptosis; mouse.
XX
OS Mus musculus.
XX
PN JP2002322082-A.
PD 08-NOV-2002.
XX
PF 26-APR-2001; 2001JP-00129200.
XX
PR 26-APR-2001; 2001JP-00129200.
XX
PA (PROT-) PROTEGENE KK.
XX
DR WPI: 2003-367092/35.
XX N-PSDB; ADC53851.
XX
XX Agents for prevention and treatment of nephritis, comprise galectin-1,
PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration
PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8
PT positive cells.
XX
XX Claim 1; SEQ ID NO 3; 31pp; Japanese.
XX
XX This invention relates to the use of novel mammal derived galectin-1
CC (G1), -3 (G3) and -9 (G9) proteins as effective ingredients for
CC prevention and treatment of nephritis. The invention discloses agents for
CC prevention and treatment of nephritis, particularly glomerular nephritis
CC and may have antinflammatory activities. The method of the invention
CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and
CC apoptosis of CD8 positive cells. The method and sequences of the
CC invention may be used for prevention and treatment of nephritis,
CC particularly glomerular nephritis including inhibition of glomerular
CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8
CC positive cells. The present sequence represents the mouse galectin 9
CC protein of the invention.
XX
XX Sequence 322 AA;
SQ
Query Match 22.0%; Score 299; DB 7; Length 322;
Best Local Similarity 37.5%; Pred. No. 1e-15;
Matches 69; Conservative 28; Mismatches 47; Indels 40; Gaps 8;
QY 63 HGAPG---AVPGAPAGVVPSPGPGAVPSSGQSPAPGAYATGPYGAAPAGPLIIVYNLP 119

```

```

DB 168 HSTPGMSTPGIP-PVYVPTP-----AYT-----IPFYTP 197
QY 120 LPGVVPMILITILGTVPKPNARIALDFORGNDVAFHF-PRENENRRVIYCNPTLDNNW 178
DB 198 IPNGIYPSKSIIMISGNVLDPATRFHINLRCCGDIAPFHNPRENEN---AVANNIQINNSW 254
QY 179 GREERQSY--FPFSGKPEFKIQVLVEPDHFKAVDNHLQ-QYNHRVKKLINEISKI 235
DB 255 GQERSLLGRPFPSGQSFVWIIICGHCFKVAVNGMCEYHYHKLXNLDINTLELVAGD 314
QY 236 IDLT 239
DB 315 IQLT 318

RESULT 49
AAV06997
ID AAV06997 standard; protein; 355 AA.
XX AAV06997;
AC
XX
XX 02-JUL-1999 (first entry)
DT
XX
DE Galectin-9 protein sequence.
KW Galectin-9 protein sequence.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO904265-A2.
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US014679.
XX
PR 17-JUL-1997; 97US-00896164.
PR 10-OCT-1997; 97US-0061599P.
PR 10-OCT-1997; 97US-0061765E.
PR 10-OCT-1997; 97US-00948705.
PR 11-OCT-1997; 97GB-00021697.
PR 22-JUN-1998; 98US-00102322.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
PI O'hare M, Obata Y, Fireundschuh M, Tureci O, Sahin U;
XX
XX WPI: 1999-132448/11.
XX N-PSDB; AAX40198.
XX
XX New isolated cancer associated nucleic acids and polypeptides - isolated
PT using sera from cancer patients, used to develop products for the
PT diagnosis, monitoring or treatment of cancers.
XX
XX Example 8; Page 779-780; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer
XX

```

SQ Sequence 355 AA;

Query Match 21.3%; Score 289; DB 2; Length 355;
Best Local Similarity 36.2%; Pred. No. 7e-15;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;

QY 66 PGAYGAPGAP-----PGVYPPGPPSG-----PGAYPSSGGP-----SAPGAYAT 102
DB 150 PRTVPQPAFSTVPFSQPCVCPFPFRGRQKPPGVWPANPAPITQTVIHVQSAFGQWFS 209
QY 103 GP-----YGAPAGPLIVPNLPLPGGVPPRMLITLTGVKPNANRLMDFORGNDAVAFH 156
DB 210 TPAIPPMYPPHAYPM--PFTTTLTGGLYPSKSTLLSGTVLPSAQRFHINLCSGNHIAFH 267
QY 157 F-PFENNRRVIVCNTKLDNNWGREERQ--SVPFESGKPEFKIQVLEPHFKAIVADA 213
DB 268 LNPFFDEN--AVVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGQ 324
QY 214 HL-QYNHRVKKLNEISKIGISGIDILT 239
DB 325 HLFEPYHRLRLPTINRLEVGDIQILT 351

RESULT 50

AAW85664

ID AAW85664 standard; protein; 355 AA.

AAW85664;

DT 19-JUL-1999 (first entry)

DE Galectin-9 like protein.

KW Galectin-9; lectin; galactose; Hodgkin's disease; pharmaceutical;
sugar chain; intercellular adhesion; cell proliferation.

OS Homo sapiens.

FN WO9910490-A1.

PD 04-MAR-1999.

PF 19-AUG-1998; 98WO-JP003670.

PR 22-AUG-1997; 97JP-00226468.

PA (SAGA) SAGAMI CHEM RES CENTRE.

PI Kato S, Yamaguchi T, Sekine S, Kamata K;

DR N-PSDB; AAX08490, AAX08491.

PT A new human protein having lactose binding properties.

PS Claim 2; Page 55-57; 64pp; English.

CC Galectins are the general term for animal lectins binding to galactose.
CC Animal lectins exist in many sites such as the cytoplasm, the nucleus,
CC the cell membrane etc. and are considered to be associated with cell
CC proliferation. Galectin-9 has been identified as an antigenic protein
CC reacting with an antibody contained in the serum of Hodgkin's disease and
CC has a structure where two sugar chain-binding domains are connected by a
CC linker peptide. The true role of galectin-9 in the body has not been
CC completely identified but is thought to be involved in intercellular
CC adhesion. The human galectin-9 like protein coding sequences are
CC characterised by containing the sequence described in AAX08489. The
CC protein can be used as pharmaceuticals or reagents for sugar chain
CC research. The cDNA is used as a probe for gene diagnosis and for gene
CC therapy

SQ Sequence 355 AA;

Query Match 21.3%; Score 289; DB 2; Length 355;
Best Local Similarity 36.2%; Pred. No. 7e-15;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;

QY 66 PGAYGAPGAP-----PGVYPPGPPSG-----PGAYPSSGGP-----SAPGAYAT 102
DB 150 PRTVPQPAFSTVPFSQPCVCPFPFRGRQKPPGVWPANPAPITQTVIHVQSAFGQWFS 209
QY 103 GP-----YGAPAGPLIVPNLPLPGGVPPRMLITLTGVKPNANRLMDFORGNDAVAFH 156
DB 210 TPAIPPMYPPHAYPM--PFTTTLTGGLYPSKSTLLSGTVLPSAQRFHINLCSGNHIAFH 267
QY 157 F-PFENNRRVIVCNTKLDNNWGREERQ--SVPFESGKPEFKIQVLEPHFKAIVADA 213
DB 268 LNPFFDEN--AVVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGQ 324
QY 214 HL-QYNHRVKKLNEISKIGISGIDILT 239
DB 325 HLFEPYHRLRLPTINRLEVGDIQILT 351

Search completed: August 23, 2004, 14:37:21
Job time : 69 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:38:48 ; Search time 123 Seconds

(without alignments)
628.510 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357

Sequence: 1 MADNPSLHDLGSGNPNPQ.....ISKLGISGIDITLSASYTWI 246

Scoring table: BLOSUM62

Searched: 1295152 segs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US09A_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1315	96.9	250	9	US-09-263-689-10
2	1315	96.9	250	9	US-09-981-353-127
3	1315	96.9	250	10	US-09-877-790-2
4	1315	96.9	250	10	US-09-919-039-298
5	1315	96.9	250	12	US-10-133-234A-1
6	1315	96.9	250	12	US-10-398-519-11
7	1315	96.9	250	14	US-10-235-674-10
8	1315	96.9	250	14	US-10-153-668-346
9	1315	96.9	250	16	US-10-409-786-3
10	1305	96.2	250	12	US-10-398-519-10
11	1297	95.6	250	12	US-10-398-519-6
12	1297	95.6	250	12	US-10-398-519-8
13	1297	95.6	250	12	US-10-398-519-12
14	1297	95.6	277	14	US-10-106-698-5787
15	1079	79.5	262	9	US-09-263-689-14

16	1079	79.5	262	12	US-10-398-519-13	Sequence 13, Appl
17	1079	79.5	262	14	US-10-235-674-14	Sequence 14, Appl
18	1064.5	78.4	263	12	US-10-398-519-14	Sequence 14, Appl
19	1049.5	77.3	245	12	US-10-133-234A-6	Sequence 6, Appl
20	1024	75.5	242	12	US-10-133-234A-4	Sequence 4, Appl
21	752	55.4	149	9	US-09-877-790-6	Sequence 6, Appl
22	723	53.3	143	10	US-10-409-786-1	Sequence 1, Appl
23	723	53.3	143	16	US-10-409-786-2	Sequence 1, Appl
24	723	53.3	144	16	US-10-409-786-2	Sequence 2, Appl
25	695.5	51.0	262	12	US-10-133-234A-5	Sequence 5, Appl
26	299	22.0	353	12	US-10-633-035-8	Sequence 8, Appl
27	295	21.7	322	9	US-09-728-479-11	Sequence 11, Appl
28	289	21.3	355	12	US-10-415-566-1	Sequence 1, Appl
29	289	21.3	355	12	US-10-633-035-2	Sequence 2, Appl
30	289	21.3	355	12	US-10-633-035-6	Sequence 6, Appl
31	289	21.3	355	14	US-10-376-133-18	Sequence 18, Appl
32	289	21.3	378	9	US-09-738-973-439	Sequence 439, App
33	289	21.3	378	9	US-09-854-133-439	Sequence 439, App
34	289	21.3	378	14	US-10-144-649A-439	Sequence 439, App
35	288	21.2	139	12	US-10-133-234A-3	Sequence 3, Appl
36	283	20.9	323	9	US-09-728-479-2	Sequence 2, Appl
37	283	20.9	323	12	US-10-415-566-2	Sequence 2, Appl
38	283	20.9	323	14	US-10-024-298A-141	Sequence 141, App
39	283	20.9	323	14	US-10-042-211A-141	Sequence 141, App
40	283	20.9	323	16	US-10-617-217A-141	Sequence 141, App
41	271	20.0	355	15	US-10-138-588-76	Sequence 76, Appl
42	267	19.7	322	12	US-10-633-035-7	Sequence 7, Appl
43	267	19.7	323	9	US-09-728-479-12	Sequence 12, Appl
44	264	19.5	311	9	US-09-263-689-4	Sequence 4, Appl
45	264	19.5	311	12	US-10-415-566-3	Sequence 3, Appl
46	264	19.5	311	14	US-10-235-674-8	Sequence 8, Appl
47	250.5	18.5	145	9	US-09-728-479-8	Sequence 8, Appl
48	250.5	18.5	145	9	US-09-894-526-5	Sequence 5, Appl
49	250.5	18.5	145	9	US-09-894-526-5	Sequence 5, Appl
50	250.5	18.5	145	14	US-10-235-674-12	Sequence 12, Appl
51	250.5	18.5	145	14	US-10-156-136-42	Sequence 42, Appl
52	250.5	18.5	145	15	US-10-455-366-3	Sequence 3, Appl
53	249	18.3	149	9	US-09-894-526-3	Sequence 3, Appl
54	246.5	18.2	168	9	US-09-922-217-139	Sequence 139, App
55	246.5	18.2	168	9	US-09-833-263-199	Sequence 199, App
56	246.5	18.2	168	13	US-10-025-380-199	Sequence 199, App
57	243.5	17.9	145	9	US-09-894-526-1	Sequence 1, Appl
58	240.5	17.7	246	15	US-10-138-588-78	Sequence 78, Appl
59	236.5	17.4	324	9	US-09-728-479-7	Sequence 7, Appl
60	236.5	17.4	324	9	US-09-263-689-11	Sequence 11, Appl
61	236.5	17.4	324	14	US-10-235-674-11	Sequence 11, Appl
62	228.5	16.8	359	14	US-10-376-133-20	Sequence 20, Appl
63	225.5	16.6	351	14	US-10-106-698-4502	Sequence 4502, Ap
64	223.5	16.5	183	10	US-09-948-227-8	Sequence 8, Appl
65	223.5	16.5	200	9	US-09-263-689-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-09-263-689-10
Sequence 10, Application US/09263689
Patent No. US20020150970A1
GENERAL INFORMATION:
APPLICANT: NI, Jiam
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 105V
NUMBER OF SEQUENCE ADDRESSES: 60
CORRESPONDENCE ADDRESSES:
ADDRESS: Steine, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D C
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,689
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/946,914
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-263-689-10

Query Match 96.9%; Score 1315; DB 9; Length 250;
Best Local Similarity 98.4%; Pred. No. 9.8e-87;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNPNPGWPGAMGNQPAAGGYPGASYPG-YPGQADPPGAYPGQAPP 59
DB 1 MAANFSLHDALSGSGNPNPGWPGAMGNQPAAGGYPGASYPGAYPGQADPPGAYPGQAPP 60
QY 60 GAYHGAAPGAYPAPAPGVYPPSPGAYSSGQPSAPGAY-ATGPYGAAPGLIIVPNYL 118
DB 61 GAYHGAAPGAYPAPAPGVYPPSPGAYSSGQPSAPGAYPATGPYGAAPGLIIVPNYL 120
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHF-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHFPRFENNRRVIVCNTKLDNN 180
QY 178 WGRERQSVFPFESGKPFKIQLVLEPDHFKAIVANDAH-LOYNHRVKKLNEISKLGISGDI 236
DB 181 WGRERQSVFPFESGKPFKIQLVLEPDHFKAIVANDAHLOYNHRVKKLNEISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 2
US-09-981-353-127
Sequence 127, Application US/09981353
Patent No. US20020160382A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 127
LENGTH: 250
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 898779CDD1
US-09-981-353-127

Query Match 96.9%; Score 1315; DB 9; Length 250;
Best Local Similarity 98.4%; Pred. No. 9.8e-87;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNPNPGWPGAMGNQPAAGGYPGASYPG-YPGQADPPGAYPGQAPP 59
DB 1 MADNFSIHDALSGSGNPNPGWPGAMGNQPAAGGYPGASYPGAYPGQADPPGAYPGQAPP 60
QY 60 GAYHGAAPGAYPAPAPGVYPPSPGAYSSGQPSAPGAY-ATGPYGAAPGLIIVPNYL 118
DB 61 GAYHGAAPGAYPAPAPGVYPPSPGAYSSGQPSAPGAYPATGPYGAAPGLIIVPNYL 120
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHF-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHFPRFENNRRVIVCNTKLDNN 180
QY 178 WGRERQSVFPFESGKPFKIQLVLEPDHFKAIVANDAH-LOYNHRVKKLNEISKLGISGDI 236
DB 181 WGRERQSVFPFESGKPFKIQLVLEPDHFKAIVANDAHLOYNHRVKKLNEISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 3
US-09-877-790-2
Sequence 2, Application US/09877790
Publication No. US20030054982A1

GENERAL INFORMATION:
APPLICANT: Jarvis, Gary
APPLICANT: Leffler, Hakon
TITLE OF INVENTION: N-TERMINALLY TRUNCATED GALECTIN-3 FOR USE IN TREATING CANCER
FILE REFERENCE: 3157.00004
CURRENT APPLICATION NUMBER: US/09/877,790
CURRENT FILING DATE: 2001-06-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 250
TYPE: PRT
ORGANISM: homo sapien
US-09-877-790-2

Query Match 96.9%; Score 1315; DB 10; Length 250;
Best Local Similarity 98.4%; Pred. No. 9.8e-87;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNPNPGWPGAMGNQPAAGGYPGASYPG-YPGQADPPGAYPGQAPP 59
DB 1 MADNFSIHDALSGSGNPNPGWPGAMGNQPAAGGYPGASYPGAYPGQADPPGAYPGQAPP 60
QY 60 GAYHGAAPGAYPAPAPGVYPPSPGAYSSGQPSAPGAY-ATGPYGAAPGLIIVPNYL 118
DB 61 GAYHGAAPGAYPAPAPGVYPPSPGAYSSGQPSAPGAYPATGPYGAAPGLIIVPNYL 120
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHF-PRFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHFPRFENNRRVIVCNTKLDNN 180
QY 178 WGRERQSVFPFESGKPFKIQLVLEPDHFKAIVANDAH-LOYNHRVKKLNEISKLGISGDI 236
DB 181 WGRERQSVFPFESGKPFKIQLVLEPDHFKAIVANDAHLOYNHRVKKLNEISKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 4
US-09-919-039-298

; Sequence 298, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaefer, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919, 039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 298
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 898779GCD1
US-09-919-039-298

Query Match 96.9%; Score 1315; DB 10; Length 250;
Best Local Similarity 98.4%; Pred. No. 9, 8e-87;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59
DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPGAYPGQAPPGAYPGQAP 60
QY 60 GAYHGAPGAYPGAPGAYPGGPGSGGAYPSSGQPSAPGAY-ATGPGYGAAPGLIYPYNTL 118
DB 61 GAYHGAPGAYPGAPGAYPGGPGSGGAYPSSGQPSAPGAYPATGPGYGAAPGLIYPYNTL 120
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHF-PPFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHFNNRRVIVCNTKLDNN 180
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKAANDAH-IQYNHRVKKLNESKLGISGDI 236
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKAANDAHILQYNHRVKKLNESKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 5
US-10-133-234A-1
; Sequence 1, Application US/10133234A
; Publication No. US20040071684A1
; GENERAL INFORMATION:
; APPLICANT: Panjwani et al.
; TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the
; FILE REFERENCE: 2002458-0006
; CURRENT APPLICATION NUMBER: US/10/133,234A
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Human
US-10-133-234A-1

Query Match 96.9%; Score 1315; DB 12; Length 250;
Best Local Similarity 98.4%; Pred. No. 9, 8e-87;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59
DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPGAYPGQAPPGAYPGQAP 60

QY 60 GAYHGAPGAYPGAPGAYPGGPGSGGAYPSSGQPSAPGAY-ATGPGYGAAPGLIYPYNTL 118
DB 61 GAYHGAPGAYPGAPGAYPGGPGSGGAYPSSGQPSAPGAYPATGPGYGAAPGLIYPYNTL 120
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHF-PPFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHFNNRRVIVCNTKLDNN 180
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKAANDAH-IQYNHRVKKLNESKLGISGDI 236
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKAANDAHILQYNHRVKKLNESKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 6
US-10-398-519-11
; Sequence 11, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPRENOL-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THREEOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-398-519-11

Query Match 96.9%; Score 1315; DB 12; Length 250;
Best Local Similarity 98.4%; Pred. No. 9, 8e-87;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59
DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPGAYPGQAPPGAYPGQAP 60
QY 60 GAYHGAPGAYPGAPGAYPGGPGSGGAYPSSGQPSAPGAY-ATGPGYGAAPGLIYPYNTL 118
DB 61 GAYHGAPGAYPGAPGAYPGGPGSGGAYPSSGQPSAPGAYPATGPGYGAAPGLIYPYNTL 120
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHF-PPFENNRRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHFNNRRVIVCNTKLDNN 180
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKAANDAH-IQYNHRVKKLNESKLGISGDI 236
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKAANDAHILQYNHRVKKLNESKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 7
US-10-235-674-10
; Sequence 10, Application US/10235674
; Publication No. US20030040081A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560004


```

Db      121 PLPGGVPRMLITLIGTVKPNANRIALDFORGNDVAHFHFNPRNNRRVIVCNTKLDNN 180
;
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-6
Query Match      95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

```

RESULT 10
US-10-398-519-10
; Sequence 10, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPEXOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 10-04
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-10

```

```

Query Match      96.2%; Score 1305; DB 12; Length 250;
Best Local Similarity 98.0%; Pred. No. 5.1e-86;
Matches 245; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

```

```

;
; US-10-398-519-6
; Sequence 6, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPEXOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6

```

```

;
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-6

```

```

Query Match      95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

```

;
; US-10-398-519-8
; Sequence 8, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPEXOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-8

```

```

;
; US-10-398-519-8
; Sequence 8, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPEXOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-8

```

```

Query Match      95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

```

;
; US-10-398-519-6
; Sequence 6, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPEXOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6

```

Db 241 DLTSASYMTI 250

RESULT 13
US-10-398-519-12
; Sequence 12, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPRENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THREEOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-398-519-12

Query Match 95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSNPNPQGMFGAMGNOPAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59
Db 1 MADNFSLHDALSGSNPNPQGMFGAMGNOPAGAGGYPGASYPGAYPGQAPPGAYPGQAP 60
QY 60 GAYHGAAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAY-ATGPGAPAGPLIVPYNL 118
Db 61 GAYPGAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAYPATGPGYGAAPGLIVPYNL 120
QY 119 PLPGGVPRMLTTITIGTKPNANRIALDFQRGNDVAFHF-PRFNENNRVIVCNTKLDNN 177
Db 121 PLPGGVPRMLTTITIGTKPNANRIALDFQRGNDVAFHFPRFNENNRVIVCNTKLDNN 180
QY 178 WGEREQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNHRVKKNEISKLGISGDI 236
Db 181 WGEREQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKNEISKLGISGDI 240
QY 237 DLTSASYMTI 246
Db 241 DLTSASYMTI 250

RESULT 14
US-10-106-698-5787
; Sequence 5787, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,117
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5787
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE

LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (125)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5787

Query Match 95.6%; Score 1297; DB 14; Length 277;
Best Local Similarity 97.6%; Pred. No. 2.1e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSNPNPQGMFGAMGNOPAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59
Db 28 MADNFSLHDALSGSNPNPQGMFGAMGNOPAGAGGYPGASYPGAYPGQAPPGAYPGQAP 87
QY 60 GAYHGAAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAY-ATGPGAPAGPLIVPYNL 118
Db 88 GAYXGAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAYPATGPGYGAAPGLIVPYNL 147
QY 119 PLPGGVPRMLTTITIGTKPNANRIALDFQRGNDVAFHF-PRFNENNRVIVCNTKLDNN 177
Db 148 PLPGGVPRMLTTITIGTKPNANRIALDFQRGNDVAFHFPRFNENNRVIVCNTKLDNN 207
QY 178 WGEREQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNHRVKKNEISKLGISGDI 236
Db 208 WGEREQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKNEISKLGISGDI 267
QY 237 DLTSASYMTI 246
Db 268 DLTSASYMTI 277

RESULT 15
US-09-263-689-14
; Sequence 14, Application US/09263689
; Patent No. US20020150970A1
; GENERAL INFORMATION:
; APPLICANT: Nt, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,689
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/946,914
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K. 36,688
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488, 0560001/EKS/SGM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 262 amino acids
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-263-689-14

```

Query Match	79.5%;	Score 1079;	DB 9;	Length 262;
Best Local Similarity	77.9%;	Pred. No. 8.9e-70;		
Matches 205;	Conservative 16;	Mismatches 24;	Indels 18;	Gaps 7

QY 1 MAINFSLHDLASSGSGPNPQGWPGAMGNQPAAGGYPGASYPG--YPGAPPGAYPGGAP 55

Db 1 MADGFSLNDLASSGSGPNPQGWPGAMGNQPGAGYFPASTPGAYPGQAPPGYGGAPP 59

QY 60 GAHHG-----APGAYPGADPAGVYPPSPSGGAYPSS--GGPSAHGAY-ATPGY 105

Db 60 SAAPGGTGPSSAYPGPTAPAGYPGPTAPAGAPFPGQPPGPGAYPSAAGAYPSAAGAYPATGPF 119

QY 106 GAPAGGLIAPVNYPLETGGGVVPPRMILTTIIGTVKPNPNRILALDFQRGNVAAHF--PREFENN 164

Db 120 GAPGTGLIYPIYDMPLPGVYMPKMLITTIIGYKPNANSLITLNFKKGNDAIRFPNRFENN 179

QY 165 RRYVCNLYKLNNWNGEERQSVFPESGKPKFIQVLVEPDHFKYAVNDAA-LQYNHVKK 223

Db 180 RRYVCNLTQDNNWNGEERQSAFPESGKPKFIQVLVEADHFYAVNDVHLQYNHRMKY 229

QY 224 LNFISLGLSGSDIDLTSASVTMI 246

Db 240 LRFISQLGIIGDITLTLSASHMT 262

```

RESULT 16
US-10-398-519-13
; Sequence 13, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPRENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 262
; TYPE: prt
; ORGANISM: Rattus sp.
; US-10-398-519-13

```

Query Match	79.5%	Score 1079;	DB 12;	Length 262;
Best Local Similarity	77.9%	Pred NC	8.9e-70.	

Matches	205; Conservative	16; Mismatches	24; Indels	19; Gaps	7
QY	1	MADNLSLDDALSGGSENPFGQWPGAKMNGPAGAGCTGTGASYEG-YEGCAPPGAYPGQAP	59		
Db	1	MADGSLNDLALGSGSENPFGQWPGAKMNQPGAGGTPGASYPGQAPPGCGYPGQAPP	59		
QY	60	GAYHG-----AFGAYGAPAPGVYGGPSSGGAYGSS--DQSPAPGAY-ATGPY	105		
Db	60	SAYPGTPGSAYPGPTAPGAYPGPTAPGAYPGQDGGGGAIFSPAPGAYPSPAPGAYPTGPF	119		
QY	106	GAPAGPLVFNLPFGGVVPRMLITLTGYKKPNANRIALDPQGNQDVAFHF-PRENNEN	164		
Db	120	GAPTEPLVFPYDMLPGGVMPRMLITLTGYKKPNANSITLTFKGGNDIFHFNPRENNEN	179		
QY	165	RRVIVCNKTLNNMGREERQGVFPFESGSEPFKIQVLEPHPCVAANDAA-LQYHNRVYK	223		
Db	180	RRVIVCNKQNNMGREERQGAFFESGSEPFKIQVLEADHEVVAANDVHLQYHNRKN	239		

```

QY      224 LNEISKLGISGDIDLTSSAYTMI 246
          |||:|||||
Db      240 LREISQLGITGDIYTSASHAMI 262

```

```

RESULT 17
US-10-235-674-14
: Sequence 14, Application US/10235674
: Publication No. US20030040081A1
: GENERAL INFORMATION:
: APPLICANT: Ni, Jian
: APPLICANT: Gentz, Rainer L.
: APPLICANT: Ruben, Steven M.
: TITLE OF INVENTION: Galectin 9 and 105V Polynucleotides
: FILE REFERENCE: 1486.0560004
: CURRENT APPLICATION NUMBER: US/10/235, 674
: CURRENT FILING DATE: 2002-09-06
: PRIOR APPLICATION NUMBER: US 09/656,450
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: US 09/263,689
: PRIOR FILING DATE: 1999-03-05
: PRIOR APPLICATION NUMBER: US 08/946,914
: PRIOR FILING DATE: 1997-10-09
: PRIOR APPLICATION NUMBER: US 60/028,093
: PRIOR FILING DATE: 1996-10-09
: NUMBER OF SEQ ID NOS: 60
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 14
: LENGTH: 262
: TYPE: PRT
: ORGANSIM: Rat
US-10-235-674-14

```

Query Match	79.5%;	Score 1079;	DB 14;	Length 262;
Best local Similarity	77.9%;	Pred. No. 8.9e-70;		
Matches 205;	Conservative 16;	Mismatches 24;	Indels 18;	Gaps 7;

QY	1	MADNLSLHDAISGSSGNPNPDGCPGAGMGKQCPGAGAGYPSAYPG-YPGQAPPGAYPGQAP	59
Db	1	MADGSLINDALAGSGNPNPDGCPGAGMGKQCPGAGAGYPSAYPG-YPGQAPPGAYPGQAP	59
QY	60	GAIHG-----APGAYPGAPAPVYDGPSSGPGAYPS--GGPSAPGAY-ATGPY	105
Db	60	SAYPPTGTSAPAYGPTATGAYAGYGPAPTAPGAPFGCGGPGAYPSAPGAYPSAPGAYPATGPF	119
QY	106	GAPAPGLIAYPYLPLPGGVVPRMLITITIGTYPKANLIALDPOKGNVAEHH-PRFENN	164
Db	120	GAPTEPLIYVYDMPLPGGVPRMLITITIGTYPKANSLITLNFKGNLIAHFNPRFENN	179
QY	165	RRVIVCNTRLDNNGGREGROSVFPFESGKPEKIVLVEPDHFKYAVNDAA-LOYNHVKK	223
Db	180	RRVIVCNTRKDNNGGREGROSGAPPFESGKPEKIVLVEADHFKYAVNDVHLLQYNHRMKV	239
QY	224	INEISKLIGSDIDLTSSATMI	246
Db	240	IREISQLIGIDITLTSSAHMI	262

```

RESULT 18
US-10-398-519-14
; Sequence 14, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPEPTID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THREEOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398, 519
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237, 858
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ. ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1

```

SEQ ID NO 14
LENGTH: 263
TYPE: PRT
ORGANISM: Mus sp.
US-10-398-519-14

Query Match 78.4%; Score 1064.5; DB 12; Length 263;
Best Local Similarity 77.6%; Pred. No. 9,8e-69;
Matches 208; Conservative 16; Mismatches 17; Indels 27; Gaps 9;

QY 1 MADNFSIHDALSGSGNPNPQGMWPGAWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADSEFLNHALSGSGNPNPQGMWPGAWGNOP-GAGGYPGAYPGAYPGQAPP 59
QY 60 GAYHG-----APGAYPGAPAPGYPPGPGGAYPPSSGQPSAGAY----- 100
DB 60 GAYPGQAPPSAYPGPTAPGAYPGPTAPGAYPGQPA-PGAFPP-GQPGAGAYPGQSGGYP 116
QY 101 ATGPYCAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDQKNDVAFHF-PR 159
DB 117 AAGP-GVPAGPLIVPYDLPLPGGVPRMLITLIGTVKPNANRIALDQKNDVAFHFPR 175
QY 160 FNNENRIVICNTKLNMMWGEREQSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQYN 218
DB 176 FNNENRIVICNTKODNNWGEREQSAFPESGKPFKIQVLVEADHFKVAVNDAAHLLQYN 235
QY 219 HRVKKLNEISKLGISGDIDITLSASYTMI 246
DB 236 HRMKNLREISQIGISGDITLSANHAM 263

RESULT 19
US-10-133-234A-6
Sequence 6, Application US/10133234A
Publication No. US20040071684A1
GENERAL INFORMATION:
APPLICANT: Panjwani et al.
TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the
FILE REFERENCE: 2002458-0006
CURRENT APPLICATION NUMBER: US/10/133,234A
CURRENT FILING DATE: 2002-04-26
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 245
TYPE: PRT
ORGANISM: Hamster
US-10-133-234A-6

Query Match 77.3%; Score 1049.5; DB 12; Length 245;
Best Local Similarity 81.0%; Pred. No. 1,1e-67;
Matches 204; Conservative 12; Mismatches 23; Indels 13; Gaps 8;

QY 1 MAONFSIHDALSGSGNPNPQGMWPGAWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADSEFLNHALSGSGNPNPQGMWPGAWGNOP-GAGGYPGAYPGAYPGQAPP 59
QY 60 GAYHG-APGAYPGAPAPGYPPGPGGAYPPSSGQPSAGAY-ATGPYCAPAGPLIVPY 116
DB 60 GAYPGQAPPSAYPGPTAPGAYPGPTAPGAYPGQPA-PGAFPP-GQPGAGAYPGQSGGYP 116
QY 117 NPLPGGVPRMLITLIGTVKPNANRIALDQKNDVAFHF-PRFNNENRIVICNTKLD 175
DB 114 KLPLAGGVPRMLITLIGTVKPNANRIALDQKNDVAFHFPRFNNENRIVICNTKLD 173
QY 176 NNMGEREQSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNEISKLGISG 234
DB 174 NNMGEREQSAFPESGKPFKIQVLVEADHFKVAVNDAAHLLQYNHRMKNLREINQMSISG 233
QY 235 DIDITLSASYTMI 246
DB 234 DITLSAAPTMI 245

RESULT 20
US-10-133-234A-4
Sequence 4, Application US/10133234A
Publication No. US20040071684A1
GENERAL INFORMATION:
APPLICANT: Panjwani et al.
TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the
FILE REFERENCE: 2002458-0006
CURRENT APPLICATION NUMBER: US/10/133,234A
CURRENT FILING DATE: 2002-04-26
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 242
TYPE: PRT
ORGANISM: Rabbit
US-10-133-234A-4

Query Match 75.5%; Score 1024; DB 12; Length 242;
Best Local Similarity 78.8%; Pred. No. 7,3e-66;
Matches 197; Conservative 15; Mismatches 26; Indels 12; Gaps 7;

QY 1 MADNFSIHDALSGSGNPNPQGMWPGAWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADSEFLNHALSGSGNPNPQGMWPGAWGNOPAGGYPGAYPGAYPGQAPP 59
QY 60 GAYHGAPGAPAPGYPPGPGGAYPPSSGQPSAGAY-ATGPYCAPAGPLIVPNL 118
DB 60 GYPYPG-PGAA-----GAYPGQPGGAYPPGQPGAGAYPGASPSAGPLPVYDL 112
QY 119 PLPGGVPRMLITLIGTVKPNANRIALDQKNDVAFHF-PRFNNENRIVICNTKLDNN 177
DB 113 PLPGGVPRMLITLIGTVKPNANRIALDQKNDVAFHFPRFNNENRIVICNTKLDNN 172
QY 178 WGEREQSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNEISKLGISGDI 236
DB 173 WGEREQSVFPESGKPFKIQVLVEPDHFKVAVNDAAHLLQYNHRMKNLREINQMSISGDI 232
QY 237 DITLSASYTMI 246
DB 233 QLTSAHAM 242

RESULT 21
US-09-728-479-6
Sequence 6, Application US/09728479
Patent No. US20020034726A1
GENERAL INFORMATION:
APPLICANT: KANEASAKI, SHIRO
APPLICANT: MATSUMOTO, RYOJI
APPLICANT: HIRASHIMA, MITSUOMI
TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR
FILE REFERENCE: 3914-2
CURRENT APPLICATION NUMBER: US/09/728,479
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: PCT/JP99/02952
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: JP 10/170698
PRIOR FILING DATE: 1998-06-02
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 149
TYPE: PRT
ORGANISM: Homo sapiens
US-09-728-479-6

Query Match 55.4%; Score 752; DB 9; Length 149;
Best Local Similarity 98.6%; Pred. No. 1,5e-46;
Matches 146; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 101 ATGPYAGAPGAPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHF-PR 159
| | | | |
Db 2 ATGPYAGAPGAPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHFNR 61
| | | | |
QY 160 FNNENRRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAH-LQYNHRVKK 218
| | | | |
Db 62 FNNENRRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAHLLQYNHRVKK 121
| | | | |
QY 219 HRVKLNEISKLIGSGDIDLTSAASYTMI 246
| | | | |
Db 122 HRVKLNEISKLIGSGDIDLTSAASYTMI 149
| | | | |

RESULT 22
US-09-877-790-1
; Sequence 1, Application US/09877790
; Publication No. US20030054982A1
; GENERAL INFORMATION:
; APPLICANT: Jarvis, Gary
; APPLICANT: John, Constance
; APPLICANT: Leffler, Hakon
; TITLE OF INVENTION: N-TERMINALLY TRUNCATED GALECTIN-3 FOR USE IN TREATING CANCER
; FILE REFERENCE: 3157.00004
; CURRENT APPLICATION NUMBER: US/09/877,790
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 143
; TYPE: PRF
; ORGANISM: homo sapien
US-09-877-790-1

Query Match 53.3%; Score 723; DB 10; Length 143;
Best Local Similarity 98.6%; Pred. No. 1,7e-44;
Matches 141; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 106 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHF-PRFNNNN 164
| | | | |
Db 1 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHFNNPRFNNNN 60
| | | | |
QY 165 RRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAH-LQYNHRVKK 223
| | | | |
Db 61 RRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAHLLQYNHRVKK 120
| | | | |
QY 224 LNEISKLIGSGDIDLTSAASYTMI 246
| | | | |
Db 121 LNEISKLIGSGDIDLTSAASYTMI 143
| | | | |

RESULT 23
US-10-409-786-1
; Sequence 1, Application US/10409786
; Publication No. US20040023855A1
; GENERAL INFORMATION:
; APPLICANT: John, Constance
; APPLICANT: Unger, Gretchen
; TITLE OF INVENTION: BIOLOGIC MODULATIONS WITH NANOPARTICLES
; FILE REFERENCE: 3157.00009
; CURRENT APPLICATION NUMBER: US/10/409,786
; PRIOR FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: 10/378,044
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 60/394,315
; PRIOR FILING DATE: 2002-07-08
; PRIOR APPLICATION NUMBER: 60/370,882
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/428,296
; PRIOR FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: 09/877,790
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: WO 02/100343

; PRIOR FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 143
; TYPE: PRF
; ORGANISM: homo sapien
US-10-409-786-1

Query Match 53.3%; Score 723; DB 16; Length 143;
Best Local Similarity 98.6%; Pred. No. 1,7e-44;
Matches 141; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 106 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHF-PRFNNNN 164
| | | | |
Db 1 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHFNNPRFNNNN 60
| | | | |
QY 165 RRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAH-LQYNHRVKK 223
| | | | |
Db 61 RRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAHLLQYNHRVKK 120
| | | | |
QY 224 LNEISKLIGSGDIDLTSAASYTMI 246
| | | | |
Db 121 LNEISKLIGSGDIDLTSAASYTMI 143
| | | | |

RESULT 24
US-10-409-786-2
; Sequence 2, Application US/10409786
; Publication No. US20040023855A1
; GENERAL INFORMATION:
; APPLICANT: John, Constance
; APPLICANT: Unger, Gretchen
; TITLE OF INVENTION: BIOLOGIC MODULATIONS WITH NANOPARTICLES
; FILE REFERENCE: 3157.00009
; CURRENT APPLICATION NUMBER: US/10/409,786
; PRIOR FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: 10/378,044
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 60/394,315
; PRIOR FILING DATE: 2002-07-08
; PRIOR APPLICATION NUMBER: 60/370,882
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/428,296
; PRIOR FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: 09/877,790
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: WO 02/100343
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 144
; TYPE: PRF
; ORGANISM: homo sapien
US-10-409-786-2

Query Match 53.3%; Score 723; DB 16; Length 144;
Best Local Similarity 98.6%; Pred. No. 1,7e-44;
Matches 141; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 106 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHF-PRFNNNN 164
| | | | |
Db 2 GAPAGPLIVPNLPLPGGVPRMLITLIGTVKPNANRIALDFQRGNDVAFHFNNPRFNNNN 61
| | | | |
QY 165 RRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAH-LQYNHRVKK 223
| | | | |
Db 62 RRVIVCNTRLNNMGREROSVFPESGKPKIQLVLPDHFKA VANDAHLLQYNHRVKK 121
| | | | |
QY 224 LNEISKLIGSGDIDLTSAASYTMI 246
| | | | |
Db 122 LNEISKLIGSGDIDLTSAASYTMI 144
| | | | |

```

RESULT 25
US-10-133-234A-5
; Sequence 5, Application US/10133234A
; Publication No. US20040071684A1
; GENERAL INFORMATION:
; APPLICANT: Panjwani et al
; TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the
; TITLE OF INVENTION: Re-Epithelialization of Wounds
; FILE REFERENCE: 2002458-0006
; CURRENT APPLICATION NUMBER: US/10/133,234A
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 262
; TYPE: prt
; ORGANISM: Chicken
; US-10-133-234A-5

```

Query Match	51.3%;	Score 695.5;	DB 12;	Length 262;
Best Local Similarity	58.8%;	Pred. No. 3e-42;		
Matches 150;	Conservative 23;	Mismatches 51;	Indels 31;	Gaps 14

QY	7	IHDALGSGS-----GAPNPGQWPGAMGNDPAAGAGYGPASYPGQAP--PGAYPGQAP	58
Dd	24	LHPQSLDALPANHGPANPPQGM-----NRPPGAF--AYGYPGAAPGAPGYPG--A	74
QY	59	PGAYHGPAGYAPGAPAPAGYIPPGPPSGAPAYSSGQSPAPGAY---ANGPYG-APAGGLIV	114
Dd	75	PGHHGPPGYPGPGP--PGYPGPGP--PGYP--GGP--PGYPGGPTPYSEAPAAPLV	127
QY	115	PYVLPPLPGGVVPMLLTTLITGVKPNANRIALDFORGNDVAFHF--PRENNNRRIYVONTK	173
Dd	128	PYDLPPLPAGLMPMLLITITGTVNSNRRSLDFKRGQDIAHHFNRRFEDHKRYIVONCSM	180
QY	174	LDNNNGHEEFQSV--FPFSGKRPKIOVLVEPDPFKYAVNDAH--LOYNHRVKKLNEISKLG	233
Dd	188	FOUNNGKEERTARPEFPEGPETPFKLOVLCEGDHFKYAVNDAHLLQFNREKKLNGITLCC	247
QY	232	ISGDIDLTSASYMTI	246
Dd	248	ISGDITLTSVLTSMI	262

```

RESULT 26
US-10-633-035-8
; Sequence 8, Application US/10633035
; Publication No. US20040068104AI
; GENERAL INFORMATION:
; APPLICANT: Seisai Kato
; APPLICANT: Yamaguchi Kimura
; APPLICANT: Shingo Sekine
; APPLICANT: Koujiu Kamata
; TITLE OF INVENTION: HUMAN GALECTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: GIN-6707CPUS
; CURRENT APPLICATION NUMBER: US/10/633, 035
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/485,951
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 9-226468
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03670
; PRIOR FILING DATE: 1998-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-633-035-8

```

Query Match	22.0%;	Score 299;	DB 12;	Length 353;
Best Local Similarity	37.5%;	Pred. No. 1.2e-13;		
Matches	69;	Conservative	28;	Mismatches 47;
			Indels	40;
			Gaps	8;

```

QY      63  HGAAG-- -AYGGAAPGVGPSPGSPGAYBSSGSGSAGAAVAGVYGAAGPLIAYINLP 118
Db      199 HSTGSGQMSTGIP -FVYTPP-----AYT-----IPFYTP 228

QY      120 LPGVVPEMLITITGYKPNANRLALDRGNDVAFHF -PRFENNRRVYQNTKLDNNW 178
Db      229 IPNGHYPSKSLMISGNVLPDATRPHINIRCGDIAFHLNPRNEN-- -AVANTQINNSW 285

QY      179 GRERQSV- -PPESGRPEKIQVLVEPDHFKYAVNDAH- -QYHRVYKLINEISKLGSGD 235
Db      286 GQERSSTLGRMPFSRQSGFSWILDEGHCFKYAVVAGHMCETYYHKLKQDINTLLEVAGD 345

QY      236 IDLT 239
Db      346 IQLT 349

```

RESULT 27
US-09-728-479-11
; Sequence 11, Application US/09728479
; Patent No. US20020034726A1
; SEQUENCE INFORMATION

Query Match	21.7%;	Score 295;	DB 9;	Length 322;
Best Local Similarity	37.0%;	Pred. No. 2e-13;		
Matches	68;	Conservative	29;	Mismatches 47;
				Indels 40;
				Gaps 8;

```

QY      63  HCAPFG-- -AYPGAPAPAGVYVPGSPSGAAPSQSGSPSAPAFAYATGPYGAAPAGPLVAYNP 119
Db      168  HSTPGCMESTPGIP-PVYVPTP-----AYT-----IPFYTP 197

QY      120  LFGVAVPEMLITILGTVPKPNANRIALDFQRGNDVAFHF-PRFENNRRVAVCTKTKIDNNK 178
Db      198  TNGTAPYKSIIMLSGNVLDPDARFHNIILNLCGGDIAPHLNPRFEN-- -AVRRATQINNSW 254

QY      179  GREERQSV--FPPESSKPKIOVLVVEPDHFKAVVNDPAHL-QVNHRYKVKNTIEISKLGISGD 235
Db      255  GGEERSLIKRMPPFVRGQSFVWLTCGHGCPKQAVNDSQHCEYIHRLLKNTLQDINTLLEVAGD 314

QY      236  IDLT 239
Db      315  IQLT 318

```

RESULT 28
US-10-415-586-1
; Sequence 1, Application US/10415586
; Publication No. US20040053346A1
; GENERAL INFORMATION:
; APPLICANT: HIRASHIMA, Mitsunori
; APPLICANT: YAMAUCHI, Akira


```

; APPLICANT: KAGESHITA, Toshiro
; APPLICANT: NAKAMURA, Takamori
; APPLICANT: NISHI, No. US20040053346A1om
; TITLE OF INVENTION: Predicting agent for a metastasis
; FILE REFERENCE: 2003-0572A/MMC/01332
; CURRENT APPLICATION NUMBER: US/10/415,586
; PRIOR FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: PCT/JP01/09561
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: JP 2000-335077
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-415-586-1
```

```

Query Match      21.3%; Score 289; DB 12; Length 355;
Best Local Similarity 36.2%; Pred. No. 6.1e-13;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;
```

```

QY 66 PGAYGAP-----PGVYPCPPSG-----PGAYSSGQP-----SAPGAYAT 102
DB 150 PRTVVGAPASTVPFSQVPCPPRRGRGRQKPPGVWPNAPAITQTVIHTVQSAPGQMS 209
QY 103 GP-----YGAPAGLIVPYNLPLPGVVPRLITITGVKPNRRIALDFORGNDVAFH 156
DB 210 TPAIPPMYHPHAYWM--PFTITLIGLYPSKSIILSTGVLPsAQRFHINLCSGNHIAFH 267
QY 157 F-PRENENRRVIVCNKTLNNWGERQ--SVPEESGKPKIQVLVEPDHFKVAVNDA 213
DB 268 LNPRFDEN--AVVANTQIDNSWGSSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDQ 324
QY 214 HL-QYNHRVKKLNEISKIGSIDIDL 239
DB 325 HLFETYHRLNLPITINRLVGGDIQLT 351
```

```

RESULT 29
US-10-633-035-2
; Sequence 2, Application US/10633035
; Publication No. US20040068104A1
; GENERAL INFORMATION:
; APPLICANT: Seiei Kato
; APPLICANT: Yamaguchi Kimura
; APPLICANT: Shingo Sekine
; APPLICANT: Kouju Kamata
; TITLE OF INVENTION: HUMAN GALECTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
; FILE REFERENCE: GIN-6707CPUS
; CURRENT APPLICATION NUMBER: US/10/633,035
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/485,951
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 9-226468
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03670
; PRIOR FILING DATE: 1998-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-633-035-2
```

```

Query Match      21.3%; Score 289; DB 12; Length 355;
Best Local Similarity 36.2%; Pred. No. 6.1e-13;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;
QY 66 PGAYGAP-----PGVYPCPPSG-----PGAYSSGQP-----SAPGAYAT 102
```

```

DB 150 PRTVVGAPASTVPFSQVPCPPRRGRGRQKPPGVWPNAPAITQTVIHTVQSAPGQMS 209
QY 103 GP-----YGAPAGLIVPYNLPLPGVVPRLITITGVKPNRRIALDFORGNDVAFH 156
DB 210 TPAIPPMYHPHAYWM--PFTITLIGLYPSKSIILSTGVLPsAQRFHINLCSGNHIAFH 267
QY 157 F-PRENENRRVIVCNKTLNNWGERQ--SVPEESGKPKIQVLVEPDHFKVAVNDA 213
DB 268 LNPRFDEN--AVVANTQIDNSWGSSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDQ 324
QY 214 HL-QYNHRVKKLNEISKIGSIDIDL 239
DB 325 HLFETYHRLNLPITINRLVGGDIQLT 351
```

```

RESULT 30
US-10-633-035-6
; Sequence 6, Application US/10633035
; Publication No. US20040068104A1
; GENERAL INFORMATION:
; APPLICANT: Seiei Kato
; APPLICANT: Yamaguchi Kimura
; APPLICANT: Shingo Sekine
; APPLICANT: Kouju Kamata
; TITLE OF INVENTION: HUMAN GALECTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
; FILE REFERENCE: GIN-6707CPUS
; CURRENT APPLICATION NUMBER: US/10/633,035
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/485,951
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 9-226468
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03670
; PRIOR FILING DATE: 1998-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-633-035-6
```

```

Query Match      21.3%; Score 289; DB 12; Length 355;
Best Local Similarity 36.2%; Pred. No. 6.1e-13;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;
```

```

QY 66 PGAYGAP-----PGVYPCPPSG-----PGAYSSGQP-----SAPGAYAT 102
DB 150 PRTVVGAPASTVPFSQVPCPPRRGRGRQKPPGVWPNAPAITQTVIHTVQSAPGQMS 209
QY 103 GP-----YGAPAGLIVPYNLPLPGVVPRLITITGVKPNRRIALDFORGNDVAFH 156
DB 210 TPAIPPMYHPHAYWM--PFTITLIGLYPSKSIILSTGVLPsAQRFHINLCSGNHIAFH 267
QY 157 F-PRENENRRVIVCNKTLNNWGERQ--SVPEESGKPKIQVLVEPDHFKVAVNDA 213
DB 268 LNPRFDEN--AVVANTQIDNSWGSSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDQ 324
QY 214 HL-QYNHRVKKLNEISKIGSIDIDL 239
DB 325 HLFETYHRLNLPITINRLVGGDIQLT 351
```

```

RESULT 31
US-10-376-133-18
; Sequence 18, Application US/10376133
; Publication No. US20030165965A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: LGALS AS MODIFIERS OF THE CHK PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX03-014C
```


LENGTH: 378
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-144-649A-439

Query Match
 Best Local Similarity 21.3%; Score 289; DB 14; Length 378;
 36.2%; Pred. No. 6.5e-13;

Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;

QY 66 PGAYPGAPDA-----PGVYGPSPSG-----PGAYPSSGCP-----SARGAYAT 102
 DB 173 PRIVVQAFSTVPSGVCPPRRGRGRKPPGWPANPAPITQTVHTVQSAFGQMP 232
 QY 103 GP-----YGAPAGPLIVPNLPLPGVVPRLITITIGTKPNANRIALDFQGNDAFH 156
 DB 223 TPALPMPMPHAPYPM--PFITITIGLSPKSIILSGTVLPSAQRFHINCSGNHIAFH 290
 QY 157 F-PFENNRRVIVCNTKLDNNWGREERQ--SVPFESGKPKIOVLVEPDHFKAVANDA 213
 DB 291 LNPRDEN--AVANNTQIDNSWGSSEERSLPRKMPFVGQSFVWLCEAHCLKAVADVQ 347
 QY 214 HL-QYNHRVKKLINEISKLIGSDIDL 239
 DB 348 HLFEPYHRLNLPITNRLLEVGGDIQT 374

RESULT 35
 US-10-133-234A-3
 Sequence 3, Application US/10133234A
 Publication No. US20040071684A1
 GENERAL INFORMATION:
 APPLICANT: Panjwani et al.
 TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the
 TITLE OF INVENTION: Re-Epithelialization of Wounds
 FILE REFERENCE: 2002458-0006
 CURRENT APPLICATION NUMBER: US/10/133,234A
 CURRENT FILING DATE: 2002-04-26
 NUMBER OF SEQ ID NOS: 28
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 3
 LENGTH: 139
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence:
 OTHER INFORMATION: Galactoside-binding lectin domain, PF00337.
 US-10-133-234A-3

Query Match
 Best Local Similarity 21.2%; Score 288; DB 12; Length 139;
 45.1%; Pred. No. 2.7e-13;
 Matches 65; Conservative 25; Mismatches 38; Indels 16; Gaps 8;

QY 111 PLIVPNLPLPGVVPRLITITIGTKP--NANRIALDFQGNDAFH--PRENE-- 162
 DB 1 PGVYALND-----GLKPGKTLIVKGTIVAPKNAKRAVAVNGSKSEENDLVLFENPREBAH 56
 QY 163 NNRRVIVCNTKL--DNNWGREERQSVPFESGKPKIOVLVEPDHFKAVANDA-LQYNE 219
 DB 57 GDQNTVVCNSKENDNEMGTQREAAFPFGQGFBEISISVEEDKFKVKNVDGHEFFRPH 116
 QY 220 RVKLTNEISKLIGSDIDLTSASY 243
 DB 117 RL-KLEAVQYIIGIKGDIKLTISK 139

RESULT 36
 US-09-728-479-2
 Sequence 2, Application US/09728479
 Patent No. US20020034726A1
 GENERAL INFORMATION:
 APPLICANT: KANEASAKI, SHIRO
 APPLICANT: MATSUMOTO, RYOJI
 APPLICANT: HIRASHIMA, MITSUOMI

TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR
 FILE REFERENCE: 3914-2
 CURRENT APPLICATION NUMBER: US/09/728,479
 CURRENT FILING DATE: 2001-08-16
 PRIOR APPLICATION NUMBER: PCT/JP99/02952
 PRIOR FILING DATE: 1999-06-02
 PRIOR APPLICATION NUMBER: JP 10/170698
 PRIOR FILING DATE: 1998-06-02
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 2
 LENGTH: 323
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-728-479-2

Query Match
 Best Local Similarity 20.9%; Score 283; DB 9; Length 323;
 34.1%; Pred. No. 1.5e-12;
 Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPGPGAPPGAYPGAP-----GAYHAPGAYPGAPA--PGVYGPSPSGAYP 89
 DB 138 GSVGLSTISFQPPGWPANPAPITQTVHTVQSAFGQMPFALPMPMPH-----AYP 192
 QY 90 SSGQPSAPGAYATGPGAPAGPLIVPNLPLPGVVPRLITITIGTKPNANRIALDFQ 149
 DB 193 -----MPFITITIGLSPKSIILSGTVLPSAQRFHINCS 228
 QY 150 GNDVAFHF--PFENNRRVIVCNTKLDNNWGREERQ--SVPFESGKPKIOVLVEPDHF 206
 DB 229 GNHIAFHLNPRDEN--AVANNTQIDNSWGSSEERSLPRKMPFVGQSFVWLCEAHCL 285
 QY 207 KVAVNDALH-QYNHRVKKLINEISKLIGSDIDL 239
 DB 286 KVAVDGQHLFEYHRLNLPITNRLLEVGGDIQT 319

RESULT 37
 US-10-415-586-2
 Sequence 2, Application US/10415586
 Publication No. US2004005346A1
 GENERAL INFORMATION:
 APPLICANT: HIRASHIMA, Mitsuomi
 APPLICANT: YAMAUCHI, Akira
 APPLICANT: KAGESHITA, Toshiro
 APPLICANT: NAKAMURA, Takamori
 APPLICANT: NISHI, No. US2004005346A1omu
 TITLE OF INVENTION: Predicting agent for a metastasis
 FILE REFERENCE: 2003-0572A/MMC/01332
 CURRENT APPLICATION NUMBER: US/10/415,586
 CURRENT FILING DATE: 2003-09-05
 PRIOR APPLICATION NUMBER: PCT/JP01/09561
 PRIOR FILING DATE: 2001-10-31
 PRIOR APPLICATION NUMBER: JP 2000-335077
 PRIOR FILING DATE: 2000-11-01
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 2
 LENGTH: 323
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-415-586-2

Query Match
 Best Local Similarity 20.9%; Score 283; DB 12; Length 323;
 34.1%; Pred. No. 1.5e-12;
 Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPGPGAPPGAYPGAP-----GAYHAPGAYPGAPA--PGVYGPSPSGAYP 89
 DB 138 GSVGLSTISFQPPGWPANPAPITQTVHTVQSAFGQMPFALPMPMPH-----AYP 192
 QY 90 SSGQPSAPGAYATGPGAPAGPLIVPNLPLPGVVPRLITITIGTKPNANRIALDFQ 149

Db 193 -----MPFTTILGILYPSKSLILSGTVLPASQRFHINCS 228
QY 150 GNDVAFHF-PRENENNRVIVCNTKLDNNWGREERQ--SVPFESGKPKIQVLVEPDHF 206
Db 229 GNHIAFHLNRPDEN---AVVANTQIDNSWGSEERSLPRKMPFVAGQSFVWLCEAHCL 285
QY 207 KVAVNDAAH-QYNHRVKKLNEISKIGSIDIDL 239
Db 286 KVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLT 319

RESULT 38
US-10-024-298A-141
; Sequence 141, Application US/10024298A
; Publication No. US20030143540A1
; GENERAL INFORMATION:
; APPLICANT: ASAH KASEI KABUSHIKI KAISHA
; APPLICANT: AKIO MATSUDA
; APPLICANT: GOICHI HONDA
; APPLICANT: SHUJI MURAMATSU
; APPLICANT: YUKIKO NAGANO
; TITLE OF INVENTION: NF-K B Activating Gene
; FILE REFERENCE: 1254-0191P
; CURRENT APPLICATION NUMBER: US/10/024,298A
; PRIOR FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: 60/314,385
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/278,641
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/258,315
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP254018/2001
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: JP0088912/2001
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: JP402288/2000
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 141
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-298A-141

Query Match 20.9%; Score 283; DB 14; Length 323;
Best Local Similarity 34.1%; Pred. No. 1.5e-12;
Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPYQGAPGAVPGQAP-----GAYHGAAPGAYGAP--PGVYPPSPSGGAYP 89
Db 138 GSVQSLYSIFQPPGWPANPAITQTVIHTVQSAFGQSFPAIPMMYPHP-----AYP 192
QY 90 SSGQSPAPGAYATGYPGAPAGPLIVPNLPLPGVVPFPMILITLGTVENNARIALDFQR 149
Db 193 -----MPFTTILGILYPSKSLILSGTVLPASQRFHINCS 228
QY 150 GNDVAFHF-PRENENNRVIVCNTKLDNNWGREERQ--SVPFESGKPKIQVLVEPDHF 206
Db 229 GNHIAFHLNRPDEN---AVVANTQIDNSWGSEERSLPRKMPFVAGQSFVWLCEAHCL 285
QY 207 KVAVNDAAH-QYNHRVKKLNEISKIGSIDIDL 239
Db 286 KVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLT 319

RESULT 39
US-10-042-211A-141
; Sequence 141, Application US/10042211A
; Publication No. US20030170719A1
; GENERAL INFORMATION:
; APPLICANT: MATSUDA, AKIO et al.
; TITLE OF INVENTION: NFKB Activating Gene

FILE REFERENCE: 1254-0192P
; CURRENT APPLICATION NUMBER: US/10/042,211A
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: JP 2000-402288
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP 2001-088912
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: JP 2001-254018
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/258,315
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/278,640
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/314,385
; PRIOR FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 141
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-211A-141

Query Match 20.9%; Score 283; DB 14; Length 323;
Best Local Similarity 34.1%; Pred. No. 1.5e-12;
Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPYQGAPGAVPGQAP-----GAYHGAAPGAYGAP--PGVYPPSPSGGAYP 89
Db 138 GSVQSLYSIFQPPGWPANPAITQTVIHTVQSAFGQSFPAIPMMYPHP-----AYP 192
QY 90 SSGQSPAPGAYATGYPGAPAGPLIVPNLPLPGVVPFPMILITLGTVENNARIALDFQR 149
Db 193 -----MPFTTILGILYPSKSLILSGTVLPASQRFHINCS 228
QY 150 GNDVAFHF-PRENENNRVIVCNTKLDNNWGREERQ--SVPFESGKPKIQVLVEPDHF 206
Db 229 GNHIAFHLNRPDEN---AVVANTQIDNSWGSEERSLPRKMPFVAGQSFVWLCEAHCL 285
QY 207 KVAVNDAAH-QYNHRVKKLNEISKIGSIDIDL 239
Db 286 KVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLT 319

RESULT 40
US-10-617-217A-141
; Sequence 141, Application US/10617217A
; Publication No. US20040081986A1
; GENERAL INFORMATION:
; APPLICANT: MATSUDA, AKIO et al.
; TITLE OF INVENTION: NF-KB ACTIVATING GENE
; FILE REFERENCE: 1254-0229P
; CURRENT APPLICATION NUMBER: US/10/617,217A
; CURRENT FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: JP 2000-402288
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP 2001-088912
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: JP 2001-254018
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/258,315
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/278,640
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/314,385
; PRIOR FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 141
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-617-217A-141

Query Match 20.9%; Score 283; DB 16; Length 323;
 Best Local Similarity 34.1%; Pred. No. 1.5e-12;
 Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPGYGQAPPAGYAPGQAPP-----GAYHGAPGAYGAPA--PGVYPPPGPGGAYP 89
 DB 138 GSVGLSTYISFQPPGWPANPAPITQVTHVQSAFGQMFSTRPALPMMYFPH-----AYP 192
 QY 90 SSGGSPAGAYATGYPGAPGLIYPYNLPLPGGVVPRMLTTIIGTYKPNANRIALDPQR 149
 DB 193 -----MPTITLIGGLYPSKSTLLSGTVLPSAQRPHINCS 228
 QY 150 GNDVAFHF-PRFNENNRVIVCNTKLDNNMGREERQ--SVPPESGKPKIQVLVEPDHF 206
 DB 229 GNHIAFHLNRFDEN---AVARNQIDNSWGSEERSLPRKMPFVRGGSFSVWILCEAHCL 285
 QY 207 KVAVND AHL-QYNHRVKKLNEISKLGISGDI DLT 239
 DB 286 KVAVDGQHLPFYHRLNRLPTINRLVEVGDIQLT 319

RESULT 41
 US-10-138-588-76
 ; Sequence 76, Application US/10138588
 ; Publication No. US20040018594A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Alsobrook et al.
 ; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACID
 ; TITLE OF INVENTION: ENCODING THE ANTIGENS, AND METHODS OF USE
 ; FILE REFERENCE: 21402-347A
 ; CURRENT APPLICATION NUMBER: US/10/138, 588
 ; PRIOR FILING DATE: 2002-05-01
 ; PRIOR APPLICATION NUMBER: 60/288,395
 ; PRIOR FILING DATE: 2001-05-03
 ; PRIOR APPLICATION NUMBER: 60/308,901
 ; PRIOR FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: 60/313,388
 ; PRIOR FILING DATE: 2001-08-17
 ; PRIOR APPLICATION NUMBER: 60/324,757
 ; PRIOR FILING DATE: 2001-09-25
 ; PRIOR APPLICATION NUMBER: 60/288,900
 ; PRIOR FILING DATE: 2001-05-04
 ; NUMBER OF SEQ ID NOS: 203
 ; SEQ ID NO 76
 ; LENGTH: 355
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-138-588-76

Query Match 20.0%; Score 271; DB 15; Length 355;
 Best Local Similarity 34.8%; Pred. No. 1.2e-11;
 Matches 72; Conservative 30; Mismatches 67; Indels 38; Gaps 9;

QY 66 PGAYGAPA-----PGVYGGPSG-----PGAYSPGQP-----SAPGAYAT 102
 DB 150 PRTPVQPAFSTVPFQPCPPRPRGRQRKPPSVRANPAPITQVTHVQSAFGQMS 209
 QY 103 GP-----YGAPAEPLIVPNLPLPGGVVPRMLTTIIGTYKPNANRIALDPQRGNDVAFH 156
 DB 210 TPAIPPMVPPHAPFM--PFTTTPGGIYPSKSTLLSGTVLPSAQRPHINCSGSHIAFH 267
 QY 157 F-PRFNENNRVIVCNTKLDNNMGREERQ--SVPPESGKPKIQVLVEPDHFVAVNDA 213
 DB 268 MNRPFDEN---AVARNQIDNSWGSEERSLPRKMPFVRGGSFSVWILCEAHCLKVAVDGQ 324
 QY 214 HL-QYNHRVKKLNEISKLGISGDI DLT 239
 DB 325 HVEFYHRLNRLPTINRLVEVGDIQLT 351

RESULT 42
 US-10-633-035-7

Sequence 7, Application US/10633035
 ; Publication No. US20040068104A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Selsi Kato
 ; APPLICANT: Yamaguchi Kimura
 ; APPLICANT: Shingo Sekine
 ; APPLICANT: Kouju Kamata
 ; TITLE OF INVENTION: HUMAN GALACTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
 ; TITLE OF INVENTION: PROTEINS
 ; FILE REFERENCE: GIN-6707CPIUS
 ; CURRENT APPLICATION NUMBER: US/10/633, 035
 ; PRIOR FILING DATE: 2003-08-04
 ; PRIOR APPLICATION NUMBER: US/09/485,951
 ; PRIOR FILING DATE: 2000-02-17
 ; PRIOR APPLICATION NUMBER: 9-226468
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: PCT/JP98/03670
 ; PRIOR FILING DATE: 1998-08-19
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 322
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-633-035-7

Query Match 19.7%; Score 267; DB 12; Length 322;
 Best Local Similarity 33.2%; Pred. No. 2.1e-11;
 Matches 71; Conservative 27; Mismatches 72; Indels 44; Gaps 8;

QY 38 GASYPGYGQAPPAGYAPGQAPP-----GAYHGAPGAYGAPA--PGVYPPPGPGGAYP 89
 DB 137 GSVGLSTYISFQPPGWPANPAPITQVTHVQSAFGQMFSTRPALPMMYFPH-----AYP 191
 QY 90 SSGGSPAGAYATGYPGAPGLIYPYNLPLPGGVVPRMLTTIIGTYKPNANRIALDPQR 149
 DB 192 -----MPTITLIGGLYPSKSTLLSGTVLPSAQRPHINCS 227
 QY 150 GNDVAFHF-PRFNENNRVIVCNTKLDNNMGREERQ--SVPPESGKPKIQVLVEPDHF 206
 DB 228 GNHIAFHLNRFDEN---AVARNQIDNSWGSEERSLPRKMPFVRGGSFSVWILCEAHCL 284
 QY 207 KVAVND AHL-QYNHRVKKLNEISKLGISGDI DLT 239
 DB 285 KVAVDGQHLPFYHRLNRLPTINRLVEVGDIQLT 318

RESULT 43
 US-09-728-479-12
 ; Sequence 12, Application US/09728479
 ; Patent No. US20020034726A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KANEGASAKI, SHIRO
 ; APPLICANT: MATSUMOTO, RYOJI
 ; APPLICANT: HIRASHIMA, MITSUOMI
 ; TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR
 ; FILE REFERENCE: 3914-2
 ; CURRENT APPLICATION NUMBER: US/09/728,479
 ; PRIOR FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: PCT/JP99/02952
 ; PRIOR FILING DATE: 1999-06-02
 ; PRIOR APPLICATION NUMBER: JP 10/170698
 ; PRIOR FILING DATE: 1998-06-02
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 12
 ; LENGTH: 323
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-728-479-12

Query Match 19.7%; Score 267; DB 9; Length 323;
 Best Local Similarity 33.2%; Pred. No. 2.1e-11;

Matches 71; Conservative 27; Mismatches 72; Indels 44; Gaps 8;

QY 38 GASYPGYGAAPPAGYPCQAPP-----GAYHAGAGAYGAPAA--PGYPPGPPSGAYP 89
 DB 138 GSVGLSYISFQPPGVPANPAPITQTIVHTVQSAGQWSTFPAIPPMYFHP-----AYP 192

QY 90 SSGQPSAAGAVATGYGAPAGLIYPNLPLPGVPPMLITITIGTVKPNRIALDFOR 149
 DB 193 -----MPFITITIGSLYPSKSTILSGTVLPSAQRFHINLCS 228

QY 150 GNDVAFHP--PENNNRRVIVCNTKLDNNWGREERO--SVPPESGKPEKIQIVLVEPDHF 206
 DB 229 GNHIAFHLNRDEN---AVVANTQIDNSWGSESLPRKMPFVAGQSPSWILGACQL 285

QY 207 KVAIVDAHL-QYNHRYKKLINEISKLGISGIDLT 239
 DB 286 KVAVDGQHLFEYHYHRLNLPITNRLVVGSDIQLT 319

RESULT 44
 US-09-263-689-4
 ; Sequence 4, Application US/09263689
 ; Patent No. US20020150970A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NI, Jian
 ; APPLICANT: Gentz, Reiner L.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
 ; NUMBER OF SEQUENCES: 60
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
 ; STREET: 1100 New York Ave., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/263,689
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/946,914
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Steffe, Eric K.
 ; REGISTRATION NUMBER: 36,688
 ; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-2600
 ; TELEFAX: 202-371-2540
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 311 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-263-689-4

Query Match 19.5%; Score 264; DB 9; Length 311;
 Best Local Similarity 36.5%; Pred. No. 3.3e-11;
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 AGAGVAGAPAA--PGVPPSGPPSGGAPGAPGAYATGTYGAPAGLIYPNLPLPG 122
 DB 159 APGQWSTFPAIPPMYFHP-----AYP-----MPFITITIG 189

QY 123 GVPFMLITITIGTVKPNRIALDFORGNDVAFHF--PENNNRRVIVCNTKLDNNWGRE 181
 DB 159 APGQWSTFPAIPPMYFHP-----AYP-----MPFITITIG 189

DB 190 GLYPSKSTILSGTVLPSAQRFHINLCSGNHIAFHLNRPDEN---AVVANTQIDNSWGSE 246

QY 182 ERQ--SVPPESGKPEKIQIVLVEPDHFVAVNDAHL-QYNHRYKKLINEISKLGISGIDL 238
 DB 247 ERSILPRKMPFVAGQSPSWILCEAHCLKAVADGQHLFEYHYHRLNLPITNRLVVGSDIQL 306

QY 239 T 239
 DB 307 T 307

RESULT 45
 US-10-415-586-3
 ; Sequence 3, Application US/10415586
 ; Publication No. US20040053346A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HIRASHIMA, Mitsumori
 ; APPLICANT: YAMAUCHI, Akira
 ; APPLICANT: KAGESHITA, Toshio
 ; APPLICANT: NAKAMURA, Takamori
 ; APPLICANT: NISHI, No. US20040053346A1omv
 ; TITLE OF INVENTION: Predicting agent for a metastasis
 ; FILE REFERENCE: 2003-0572A/WMC/01332
 ; CURRENT APPLICATION NUMBER: US/10/415,586
 ; CURRENT FILING DATE: 2003-09-05
 ; PRIOR APPLICATION NUMBER: PCT/JP01/09561
 ; PRIOR FILING DATE: 2001-10-31
 ; PRIOR APPLICATION NUMBER: JP 2000-335077
 ; PRIOR FILING DATE: 2000-11-01
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 311
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-415-586-3

Query Match 19.5%; Score 264; DB 12; Length 311;
 Best Local Similarity 36.5%; Pred. No. 3.3e-11;
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 AGAGVAGAPAA--PGVPPSGPPSGGAPGAPGAYATGTYGAPAGLIYPNLPLPG 122
 DB 159 APGQWSTFPAIPPMYFHP-----AYP-----MPFITITIG 189

QY 123 GVPFMLITITIGTVKPNRIALDFORGNDVAFHF--PENNNRRVIVCNTKLDNNWGRE 181
 DB 190 GLYPSKSTILSGTVLPSAQRFHINLCSGNHIAFHLNRPDEN---AVVANTQIDNSWGSE 246

QY 182 ERQ--SVPPESGKPEKIQIVLVEPDHFVAVNDAHL-QYNHRYKKLINEISKLGISGIDL 238
 DB 247 ERSILPRKMPFVAGQSPSWILCEAHCLKAVADGQHLFEYHYHRLNLPITNRLVVGSDIQL 306

QY 239 T 239
 DB 307 T 307

RESULT 46
 US-10-235-674-4
 ; Sequence 4, Application US/10235674
 ; Publication No. US20030040081A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NI, Jian
 ; APPLICANT: Gentz, Reiner L.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: Galectin 8 and 10SV Polynucleotides
 ; FILE REFERENCE: 1488.0560004
 ; CURRENT APPLICATION NUMBER: US/10/235,674
 ; CURRENT FILING DATE: 2002-09-06
 ; PRIOR APPLICATION NUMBER: US 09/656,450
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: US 09/263,689

```

1 /
2 /
3 / PRIOR FILING DATE: 1999-03-05
4 / PRIOR APPLICATION NUMBER: US 08/946,914
5 / PRIOR FILING DATE: 1997-10-09
6 / PRIOR APPLICATION NUMBER: US 60/028,093
7 / PRIOR FILING DATE: 1996-10-09
8 / NUMBER OF SEQ ID NOS: 60
9 / SOFTWARE: PatentIn version 3.0
10 / SEQ ID NO 4
11 / LENGTH: 311
12 / TYPE: prt
13 / ORGANISM: Homo sapiens
14 /
15 /
16 /
17 /
18 /
19 /
20 /
21 /
22 /
23 /
24 /
25 /
26 /
27 /
28 /
29 /
30 /
31 /
32 /
33 /
34 /
35 /
36 /
37 /
38 /
39 /
40 /
41 /
42 /
43 /
44 /
45 /
46 /
47 /
48 /
49 /
50 /
51 /
52 /
53 /
54 /
55 /
56 /
57 /
58 /
59 /
60 /
61 /
62 /
63 /
64 /
65 /
66 /
67 /
68 /
69 /
70 /
71 /
72 /
73 /
74 /
75 /
76 /
77 /
78 /
79 /
80 /
81 /
82 /
83 /
84 /
85 /
86 /
87 /
88 /
89 /
90 /
91 /
92 /
93 /
94 /
95 /
96 /
97 /
98 /
99 /
100 /
101 /
102 /
103 /
104 /
105 /
106 /
107 /
108 /
109 /
110 /
111 /
112 /
113 /
114 /
115 /
116 /
117 /
118 /
119 /
120 /
121 /
122 /
123 /
124 /
125 /
126 /
127 /
128 /
129 /
130 /
131 /
132 /
133 /
134 /
135 /
136 /
137 /
138 /
139 /
140 /
141 /
142 /
143 /
144 /
145 /
146 /
147 /
148 /
149 /
150 /
151 /
152 /
153 /
154 /
155 /
156 /
157 /
158 /
159 /
160 /
161 /
162 /
163 /
164 /
165 /
166 /
167 /
168 /
169 /
170 /
171 /
172 /
173 /
174 /
175 /
176 /
177 /
178 /
179 /
180 /
181 /
182 /
183 /
184 /
185 /
186 /
187 /
188 /
189 /
190 /
191 /
192 /
193 /
194 /
195 /
196 /
197 /
198 /
199 /
200 /
201 /
202 /
203 /
204 /
205 /
206 /
207 /
208 /
209 /
210 /
211 /
212 /
213 /
214 /
215 /
216 /
217 /
218 /
219 /
220 /
221 /
222 /
223 /
224 /
225 /
226 /
227 /
228 /
229 /
230 /
231 /
232 /
233 /
234 /
235 /
236 /
237 /
238 /
239 /
240 /
241 /
242 /
243 /
244 /
245 /
246 /
247 /
248 /
249 /
250 /
251 /
252 /
253 /
254 /
255 /
256 /
257 /
258 /
259 /
260 /
261 /
262 /
263 /
264 /
265 /
266 /
267 /
268 /
269 /
270 /
271 /
272 /
273 /
274 /
275 /
276 /
277 /
278 /
279 /
280 /
281 /
282 /
283 /
284 /
285 /
286 /
287 /
288 /
289 /
290 /
291 /
292 /
293 /
294 /
295 /
296 /
297 /
298 /
299 /
300 /
301 /
302 /
303 /
304 /
305 /
306 /
307 /
308 /
309 /
310 /
311 /
312 /
313 /
314 /
315 /
316 /
317 /
318 /
319 /
320 /
321 /
322 /
323 /
324 /
325 /
326 /
327 /
328 /
329 /
330 /
331 /
332 /
333 /
334 /
335 /
336 /
337 /
338 /
339 /
340 /
341 /
342 /
343 /
344 /
345 /
346 /
347 /
348 /
349 /
350 /
351 /
352 /
353 /
354 /
355 /
356 /
357 /
358 /
359 /
360 /
361 /
362 /
363 /
364 /
365 /
366 /
367 /
368 /
369 /
370 /
371 /
372 /
373 /
374 /
375 /
376 /
377 /
378 /
379 /
380 /
381 /
382 /
383 /
384 /
385 /
386 /
387 /
388 /
389 /
390 /
391 /
392 /
393 /
394 /
395 /
396 /
397 /
398 /
399 /
400 /
401 /
402 /
403 /
404 /
405 /
406 /
407 /
408 /
409 /
410 /
411 /
412 /
413 /
414 /
415 /
416 /
417 /
418 /
419 /
420 /
421 /
422 /
423 /
424 /
425 /
426 /
427 /
428 /
429 /
430 /
431 /
432 /
433 /
434 /
435 /
436 /
437 /
438 /
439 /
440 /
441 /
442 /
443 /
444 /
445 /
446 /
447 /
448 /
449 /
450 /
451 /
452 /
453 /
454 /
455 /
456 /
457 /
458 /
459 /
460 /
461 /
462 /
463 /
464 /
465 /
466 /
467 /
468 /
469 /
470 /
471 /
472 /
473 /
474 /
475 /
476 /
477 /
478 /
479 /
480 /
481 /
482 /
483 /
484 /
485 /
486 /
487 /
488 /
489 /
490 /
491 /
492 /
493 /
494 /
495 /
496 /
497 /
498 /
499 /
500 /
501 /
502 /
503 /
504 /
505 /
506 /
507 /
508 /
509 /
510 /
511 /
512 /
513 /
514 /
515 /
516 /
517 /
518 /
519 /
520 /
521 /
522 /
523 /
524 /
525 /
526 /
527 /
528 /
529 /
530 /
531 /
532 /
533 /
534 /
535 /
536 /
537 /
538 /
539 /
540 /
541 /
542 /
543 /
544 /
545 /
546 /
547 /
548 /
549 /
550 /
551 /
552 /
553 /
554 /
555 /
556 /
557 /
558 /
559 /
560 /
561 /
562 /
563 /
564 /
565 /
566 /
567 /
568 /
569 /
570 /
571 /
572 /
573 /
574 /
575 /
576 /
577 /
578 /
579 /
580 /
581 /
582 /
583 /
584 /
585 /
586 /
587 /
588 /
589 /
590 /
591 /
592 /
593 /
594 /
595 /
596 /
597 /
598 /
599 /
600 /
601 /
602 /
603 /
604 /
605 /
606 /
607 /
608 /
609 /
610 /
611 /
612 /
613 /
614 /
615 /
616 /
617 /
618 /
619 /
620 /
621 /
622 /
623 /
624 /
625 /
626 /
627 /
628 /
629 /
630 /
631 /
632 /
633 /
634 /
635 /
636 /
637 /
638 /
639 /
640 /
641 /
642 /
643 /
644 /
645 /
646 /
647 /
648 /
649 /
650 /
651 /
652 /
653 /
654 /
655 /
656 /
657 /
658 /
659 /
660 /
661 /
662 /
663 /
664 /
665 /
666 /
667 /
668 /
669 /
670 /
671 /
672 /
673 /
674 /
675 /
676 /
677 /
678 /
679 /
680 /
681 /
682 /
683 /
684 /
685 /
686 /
687 /
688 /
689 /
690 /
691 /
692 /
693 /
694 /
695 /
696 /
697 /
698 /
699 /
700 /
701 /
702 /
703 /
704 /
705 /
706 /
707 /
708 /
709 /
710 /
711 /
712 /
713 /
714 /
715 /
716 /
717 /
718 /
719 /
720 /
721 /
722 /
723 /
724 /
725 /
726 /
727 /
728 /
729 /
730 /
731 /
732 /
733 /
734 /
735 /
736 /
737 /
738 /
739 /
740 /
741 /
742 /
743 /
744 /
745 /
746 /
747 /
748 /
749 /
750 /
751 /
752 /
753 /
754 /
755 /
756 /
757 /
758 /
759 /
760 /
761 /
762 /
763 /
764 /
765 /
766 /
767 /
768 /
769 /
770 /
771 /
772 /
773 /
774 /
775 /
776 /
777 /
778 /
779 /
780 /
781 /
782 /
783 /
784 /
785 /
786 /
787 /
788 /
789 /
790 /
791 /
792 /
793 /
794 /
795 /
796 /
797 /
798 /
799 /
800 /
801 /
802 /
803 /
804 /
805 /
806 /
807 /
808 /
809 /
810 /
811 /
812 /
813 /
814 /
815 /
816
```

Query Match	19.5%	Score 264;	DB 14;	Length 311;
Best Local Similarity	36.5%	Pred. No. 3.3e-11;		
Matches	66;	Conservative	25;	Mismatches 52;
				Indels 38;
				Gaps 7

QY 65 APAYGAP- -GVVPGSPGSGVPSGQSPAGAAVATPGYAGPGLIVFENLPLPG 122
 Db 159 APQGMSTPAIPMMVPHF-----AYP-----MPTITLIG 189
 QY 123 GVTPRNLTITLGVKPNANRIALDPORGNDVAHF- -PFNNNNRVIYVCTKLDNNGRE 181
 Db 190 GLYPSKSIILSGVLPVSAQRPHINICSGNHIAFHLNPRFDEN--AVVRNTQIDNSWGE 246
 QY 182 ERQ- -SVFPESGKPPKIQIVLEPDPFKVAVDAHL- -QYNRVKVLNLEISLTGSGPIDL 238
 Db 247 ERSLLPRKMPVRQGSFVWILCEAHCLKVAVDGHLFFYHRLTNLPTLNLELVGGPIQL 306
 QY 239 T 239
 Db 307 T 307

RESULT 47
 US-09-728-479-8
 : Sequence 8, Application US/09728479
 : Patent No. US20020034726A1
 : GENERAL INFORMATION:
 : APPLICANT: KANEGASAKI, SHIRO
 : APPLICANT: MATSUMOTO, RYOJI
 : APPLICANT: HIRASHIMA, MITSUOMI
 : TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR
 : FILE REFERENCE: 3914-2
 : CURRENT APPLICATION NUMBER: US/09/728,479
 : CURRENT FILING DATE: 2001-08-16
 : PRIOR APPLICATION NUMBER: PCT/JP99/02952
 : PRIOR FILING DATE: 1999-06-02
 : PRIOR APPLICATION NUMBER: JP 10/170698
 : PRIOR FILING DATE: 1998-06-02
 : NUMBER OF SEQ ID NOS: 12
 : SOFTWARE: Patentin Ver. 2.1
 : SEQ ID NO 6
 : LENGTH: 145
 : TYPE: PRT
 : ORGANISM: Rattus sp.
 : US-09-728-479-8

Query March	18.5%	Score 250.5;	DB 9;	Length 145;
Best Local Similarity	40.4%	Pred. No. 1.4e-10;		
Matches	55;	Conservative	26;	Mismatches 46; Indels 7; Gaps 4
QY	108	PAGELIVPYNLPFGGVPRMLITLIGYKPNARIADFGQNDVAHFH-PRFENENRR	166	
Db	9	PEYNLAVPFPTSIPLNLYPSKSLVLSGWLSDPAKRFQNLNACGGDIAFHLNPRDEN---	65	
QY	167	VIVYCNLTLDNNNGREERQ--SVFPPESGKPKFIQVLVEPDHFKYAVNDAAH-QYNHRVKK	223	
Db	66	AAVRNTOINNSWGPEERSLPGSMPSRGGRFESVWILCEGHFKVAVDGHCHESTRLMN	125	
QY	224	LNESIKLTGISGIDILT	239	
Db	126	LPDINTLEVAGDIQLT	141	

RESULT 48
US-09-894-526-5
; Sequence 5, Application US/0984526
; Patent No. US20020127689A1
GENERAL INFORMATION:

TITLE OF INVENTION: NOVEL HUMAN GALECTINS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:

COMPUTER READABLE FORM:

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

```
SOFTWARE: FastSeq for Windows Version 2.0.0
```

APPLICATION NUMBER: 08/788,584

ATTORNEY/AGENT INFORMATION:

INVOICE: BILLINGS, DUCY O.
REGISTRATION NUMBER: 36.749

REFERENCE/DOCKET NUMBER: PH-0192 US
TELECOMMUNICATION INFORMATION:

; INFORMATION FOR SEQ ID NO: 5:

LENGTH: 145 amino acids

STRANDEDNESS: single

IMMEDIATE SOURCE: 11
10F0001: 11

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-894-526-5

```

Query Match	18.5%	Score 250.5	DB 9	Length 145
Best Local Similarity	40.4%	Pred. No. 1.4e-10		
Matches	55	Conservative	46	Indels 7
				Gaps 4

```

QY      108  PGGELLVPPNLLPGGVPRMLITLLGVKNANRIADFORGNDVAFFE--PRENNRR 166
Db      9    PYPNLAVPPFTSLPGLYPSKSYIVSGVVLDAKRFQINLRGGDIAFHLNPRDEN--- 65

QY      167  VIIVCNKTLNNMGREERQ--SVFPESGSKPKRIQVLEPDHFKAIVANDAH--QYNHRVK 223
Db      66  AVAVRNQINNSGPPERSLPGSMPPSRQRFESVAILCEGHCFKVAIVDGOHCEISHKLM 125

QY      224  LNEISKLGISGDIDLTL 239
Db      126  LPDINTLEVAAGDIQILT 141

```

RESULT 49
US-09-263-689-12
; Sequence 12, Application US/09263689
; Patent No. US20020150970A1

```

; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,689
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/946,914
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-263-689-12

Query Match      18.5%; Score 250.5; DB 9; Length 145;
Best Local Similarity 40.4%; Pred. No. 1.4e-10;
Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPNLPLPGGVVPRMLITLITGVKPNANRIALDFORGNDVAFHF--PRENNENNR 166
DB 9 PYPNLAVPFPTSIIPNGLYPSKSIIVISGVLSDAKRFQINLRGCGDIAFHLNRPFDEN--- 65
QY 167 VIVCNTKLDNMWGREERQ--SVFPESGKPFKIQVLVEPDHFKVAVNDAHL-QYNHRYVK 223
DB 66 AVVRNTQINNSWGPERSLPGSMPSRGQRFQSVWILCEGHCCKVAVDGOHICEYSHRLMN 125
QY 224 LNEISKLGISGDIIDL 239
DB 126 LPDINTLEVAGDIQLT 141

RESULT 50
US-10-235-674-12
; Sequence 12, Application US/10235674
; Publication No. US20030040081A1
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560004
; CURRENT APPLICATION NUMBER: US/10/235,674
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 09/656,450
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
```

```

; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; PRIOR FILING DATE: 1996-10-09
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent In Release #1.0, Version #1.30
; SEQ ID NO: 12
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Rat
; US-10-235-674-12

Query Match      18.5%; Score 250.5; DB 14; Length 145;
Best Local Similarity 40.4%; Pred. No. 1.4e-10;
Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPNLPLPGGVVPRMLITLITGVKPNANRIALDFORGNDVAFHF--PRENNENNR 166
DB 9 PYPNLAVPFPTSIIPNGLYPSKSIIVISGVLSDAKRFQINLRGCGDIAFHLNRPFDEN--- 65
QY 167 VIVCNTKLDNMWGREERQ--SVFPESGKPFKIQVLVEPDHFKVAVNDAHL-QYNHRYVK 223
DB 66 AVVRNTQINNSWGPERSLPGSMPSRGQRFQSVWILCEGHCCKVAVDGOHICEYSHRLMN 125
QY 224 LNEISKLGISGDIIDL 239
DB 126 LPDINTLEVAGDIQLT 141
```

Search completed: August 23, 2004, 14:49:53
Job time : 138 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:36:22 ; Search time 22 Seconds
(without alignments)
577.272 Million cell updates/sec

Title: US-09-297-040-4
Perfect score: 1357
Sequence: 1 MADNPSLHDLISGSGNPNPQ.....ISKLGISGIDIDITSASYTM1 246

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1315	96.9	250	3	US-08-946-914-10
2	1315	96.9	250	4	US-09-656-450-10
3	1305	96.2	250	1	US-08-562-311-2
4	1079	79.5	262	3	US-08-946-914-14
5	1079	79.5	262	4	US-09-656-450-14
6	1078	79.4	264	1	US-08-562-311-4
7	513	37.8	135	2	US-08-647-960-5
8	277.5	20.4	322	4	US-09-559-023-2
9	264	19.5	311	3	US-08-946-914-4
10	264	19.5	311	4	US-09-656-450-4
11	250.5	18.5	145	2	US-08-788-584-5
12	250.5	18.5	145	3	US-08-946-914-12
13	250.5	18.5	145	4	US-09-656-450-12
14	250.5	18.5	145	4	US-09-557-170A-3
15	249	18.3	149	2	US-08-788-584-3
16	246.5	18.2	168	4	US-09-401-064-199
17	243.5	17.9	145	2	US-08-788-584-1
18	243.5	17.9	301	4	US-09-559-023-4
19	236.5	17.4	324	3	US-08-946-914-11
20	236.5	17.4	324	4	US-09-656-450-11
21	223.5	16.5	200	3	US-08-946-914-8
22	223.5	16.5	200	4	US-09-656-450-8
23	223.5	16.5	316	3	US-09-131-648-5
24	223.5	16.5	317	4	US-08-946-914-6
25	218.5	16.1	316	2	US-09-656-450-6
26	218.5	16.1	316	2	US-08-728-521-3
27	218.5	16.1	316	2	US-08-647-960-2

28	218.5	16.1	316	3	US-08-946-914-15	Sequence 15, Appl
29	218.5	16.1	316	3	US-08-946-914-17	Sequence 17, Appl
30	218.5	16.1	316	3	US-09-212-146-3	Sequence 3, Appl
31	218.5	16.1	316	4	US-09-656-450-15	Sequence 15, Appl
32	218.5	16.1	316	4	US-09-656-450-17	Sequence 17, Appl
33	215.5	15.9	323	3	US-08-469-667-16	Sequence 16, Appl
34	215.5	15.9	323	3	US-08-946-914-2	Sequence 2, Appl
35	215.5	15.9	323	4	US-09-224-110-16	Sequence 16, Appl
36	215.5	15.9	323	4	US-09-656-450-2	Sequence 2, Appl
37	215.5	15.9	323	3	PCT-US95-07289-16	Sequence 16, Appl
38	211	15.5	136	3	US-08-946-914-13	Sequence 13, Appl
39	211	15.5	136	4	US-09-154-750A-79	Sequence 79, Appl
40	211	15.5	136	4	US-09-656-450-13	Sequence 13, Appl
41	207	15.3	318	4	US-09-557-170A-4	Sequence 4, Appl
42	200.5	14.8	504	3	US-09-219-849-3	Sequence 4, Appl
43	200.5	14.8	561	1	US-08-642-255-52	Sequence 52, Appl
44	199.5	14.7	144	1	US-08-642-255-49	Sequence 49, Appl
45	199.5	14.7	720	3	US-09-219-849-4	Sequence 4, Appl
46	199.5	14.7	777	1	US-08-642-255-53	Sequence 53, Appl
47	197.5	14.6	234	1	US-08-642-255-51	Sequence 51, Appl
48	196.5	14.5	466	3	US-08-526-136-13	Sequence 13, Appl
49	195.5	14.4	417	1	US-08-175-155-69	Sequence 69, Appl
50	195.5	14.4	417	1	US-08-477-509B-104	Sequence 104, Appl
51	195.5	14.4	417	1	US-08-642-255-102	Sequence 102, Appl
52	195.5	14.4	417	2	US-08-707-237A-76	Sequence 76, Appl
53	195.5	14.4	417	3	US-08-482-085B-104	Sequence 104, Appl
54	195.5	14.4	417	4	US-09-444-791A-104	Sequence 104, Appl
55	195.5	14.4	829	1	US-08-642-255-132	Sequence 132, Appl
56	195.5	14.4	829	1	US-08-397-633A-53	Sequence 53, Appl
57	195.5	14.4	837	1	US-08-175-155-68	Sequence 68, Appl
58	195.5	14.4	837	1	US-08-477-509B-103	Sequence 103, Appl
59	195.5	14.4	837	1	US-08-642-255-101	Sequence 101, Appl
60	195.5	14.4	837	2	US-08-707-237A-75	Sequence 75, Appl
61	195.5	14.4	837	3	US-08-482-085B-103	Sequence 103, Appl
62	195.5	14.4	837	4	US-09-444-791A-103	Sequence 103, Appl
63	195.5	14.4	897	1	US-08-397-633A-50	Sequence 50, Appl
64	192	14.1	166	4	US-09-841-334A-21	Sequence 21, Appl
65	191	14.1	160	3	US-08-542-051-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-946-914-10
Sequence 10, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:

NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-10

Query Match 96.9%; Score 1315; DB 3; Length 250;
Best Local Similarity 98.4%; Pred. No. 4,6e-114;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGANGNPAGAGGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSGSGNPNPQGMWPGANGNPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60
QY 60 GAYHAPGAYPGAPAPAGVYPPGPPSGGAYPSSGQPSAPGAY-ATGPYCAPAPLIVPYNL 118
DB 61 GAYHAPGAYPGAPAPAGVYPPGPPSGGAYPSSGQPSAPGAYPATGPYCAPAPLIVPYNL 120
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHF-PFNENNRVIVCNKTLDNN 177
DB 121 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHFNPENNRVIVCNKTLDNN 180
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAH-LQYNHRVKLNLSKLGISGDI 236
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAHLQYNHRVKLNLSKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 2
US-09-656-450-10
Sequence 10, Application US/09656450
Patent No. 6468768
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
FILE REFERENCE: 1488.0560003
CURRENT APPLICATION NUMBER: US/09/656,450
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/263,689
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: US 08/946,914
PRIOR FILING DATE: 1997-10-09
PRIOR APPLICATION NUMBER: US 60/028,093
PRIOR FILING DATE: 1996-10-09
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn version 3.0
SEQ ID NO 10
LENGTH: 250
TYPE: PRT
ORGANISM: Homo sapiens
US-09-656-450-10

Query Match 96.9%; Score 1315; DB 4; Length 250;
Best Local Similarity 98.4%; Pred. No. 4,6e-114;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGANGNPAGAGGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSGSGNPNPQGMWPGANGNPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60

QY 60 GAYHAPGAYPGAPAPAGVYPPGPPSGGAYPSSGQPSAPGAY-ATGPYCAPAPLIVPYNL 118
DB 61 GAYHAPGAYPGAPAPAGVYPPGPPSGGAYPSSGQPSAPGAYPATGPYCAPAPLIVPYNL 120
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHF-PFNENNRVIVCNKTLDNN 177
DB 121 PLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFHFNPENNRVIVCNKTLDNN 180
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAH-LQYNHRVKLNLSKLGISGDI 236
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAHLQYNHRVKLNLSKLGISGDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 3
US-08-562-311-2
Sequence 2, Application US/08562311
Patent No. 5801002
GENERAL INFORMATION:
APPLICANT: RAZ, AVRAHAM
TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY
TITLE OF INVENTION: OF METASTASIS IN A CELL SAMPLE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Dykema Gosselt
STREET: STE 505 N. Woodward
CITY: Bloomfield Hills
STATE: MI
COUNTRY: U.S.
ZIP: 48304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/562,311
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/188,225
FILING DATE:
APPLICATION NUMBER: US 07/681,242
FILING DATE: 04-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/294,249
FILING DATE: 01-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: KELLY, ROBERT L.
REGISTRATION NUMBER: 31,843
REFERENCE/DOCKET NUMBER: 61,686-
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-540-0849
TELEFAX: 810-540-0763
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-562-311-2

Query Match 96.2%; Score 1305; DB 1; Length 250;
Best Local Similarity 97.6%; Pred. No. 3,9e-113;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGANGNPAGAGGYPGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSGSGNPNPQGMWPGANGNPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60

QY	60	GAHHAGAGATPAGAPALGAVVPPSPGGATPSSGCGSAAAGAT-ARGPGAPAGPLIVLYNL	118
Db	61	GAHHAGAGATPAGAPALGAVVPPSPGGATPSSGCGSAAAGATPATGPGAPAGPLIVLYNL	120
QY	119	PLFGGVPRMLITLITGVKPNMNRIALDFQRNDVAFHF-PRFNENNRKRVLCNTKLDNN	177
Db	121	PLFGGVPRMLITLITGVKPNMNRIALDFQRNDVAFHFNPRFNENNRKRVLCNTKLDNN	180
QY	178	WGEEEROSVPPESGKPFKIQVIVEDDHKKVAVNDAAH-LQTNHRVKKLNEISKIGSDI	236
Db	181	WGEEEROSVPPESGKPFKIHLVIEDDHKKVAVNDAAHLLQTNHRVKKLNEIRKIGISDI	240
QY	237	DLTSASYTMT	246
Db	241	DLTSASYTMT	250

```

RESULT 4
US-08-946-914-14
; Sequence 14, Application US/08946914
Patent No. 6027916
; GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-14

Query Match 79.5%; Score 1079; DB 3; Length 262;
Best Local Similarity 77.9%; Pred. No. 3,4e-92;
Matches 205; Conservative 16; Mismatches 24; Indels 18; Gaps 7

QY 1 MDNFSIDALSGSGNPNGWPGAMGQOPAGAGYPPASYPG-YPGQAPPGAYPGQAPP 59
Db 1 MADGFSIDNALGSGNPNGWPGAMGQOP-GAGGYPPASYPGAYPPQAPPGGAYPGQAPP 59
QY 60 GATHG-----APGAYPGAPAPGVYVGPSPSGPQAYPSS--GGPSAPGAY-ATGPy 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db	60	SAYPGTGBSAYPGTAPGAYGGTAPGAPGPGGPGGAYPSADGAYPSAGGAYPANGPF	119
Qy	106	GAGAGELTPYNYLPLPGGVPRMLITITITGYKPNANR ALDFQNGNDVAHFH -PREVENN	164
Db	120	GATGTGLTPYDMPLEGGVPRMLITITITGYKPNANSLITLFFKKGNDIAEFHFNPRENN	179
Qy	165	RRIVYNTLDDNNMGEEHRSVFPEESGPKIQVLVEPHEFKAVUNDH -LOYNHVKK	223
Db	180	RRIVYNTKQDDNNMGEEHRSQAFPEESGPKIQVLVEADHFKVAUNDVHLLQYNHMKK	239
Qy	224	LNIEISKLGISGIDILTSAGTYMI	246
Db	240	LRISQGLIGDITLTSASHAMI	262

```

RESULT 5
US-09-656-450-14
: Sequence 14, Application US/09656450
: Patent No. 6468768
: GENERAL INFORMATION:
: APPLICANT: Ni, Jian
: APPLICANT: Gentz, Reiner L.
: APPLICANT: Ruben, Steven M.
: TITLE OF INVENTION: Galectin 9 and 10sv Polynucleotides
: FILE REFERENCE: 1488.0560003
: CURRENT APPLICATION NUMBER: US/09/656,450
: PRIORITY FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: US 09/263,689
: PRIOR FILING DATE: 1999-03-05
: PRIOR APPLICATION NUMBER: US 08/946,914
: PRIOR FILING DATE: 1997-10-09
: PRIOR APPLICATION NUMBER: US 60/028,093
: PRIOR FILING DATE: 1996-10-09
: NUMBER OF SEQ ID NOS: 60
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 14
: LENGTH: 262
: TYPE: PRT
: ORGANISM: Rat
US-09-656-450-14

Query Match      79.5%; Score 1079; DB 4; Length 262;
Best Local Similarity 77.9%; Pred. No. 3.4e-92;
Matches 205; Conservative 16; Mismatches 24; Indels 18; Gaps 7

QY      1  MADNFSLHDALSGNGNPNPQGMFGAMGNQPGAGGYPGASYPG--YPGGAPGAYPGGAGP 59
Db      1  MADGSLNDALAGSGNRPFGCMFGAMGNQF--GAGGYPGASIFGAYFGGAYFGGQAPP 59

QY      60  GAYHG-----APGAYGAPADGVDPGPPSGPGAYSS--GQSPAPGAY-ATGPY 105
Db      60  SAYGPGTPGASYPGPTAPGAPGPTAGGAFPGQPGPGAYPSAGGAYPSAGAYPATGPF 119

QY      106  GAGPGPLIVPNNPLPGGVDPBRLITLIGTVKPMANRLALDPFGQNDVAEHF--PRENNN 164
Db      120  GATGPEPLIVEYDMPLPGGVNPRMLITLIGTVKPMANSTLTLNFKKGNDIAHFNNPRENN 179

QY      165  RRVIVCNTKLDNMNGREERQSVPFESGKPKIQLVVEPDHFKYAVNDAA--LQYNHRVKK 223
Db      180  RRVIVCNTKQDNNNGREERQSAFPFESGKPKIQLVVEADHFKYAVNDVHLQYNHRMKN 239

QY      224  LNEISKLGISGIDILTSASYMTI 246
Db      240  LREISGLIGTIDITLTSASHMTI 262

RESULT 6
US-08-562-311-4
: Sequence 4, Application US/08562311
: Patent No. 5801002
: GENERAL INFORMATION:
: APPLICANT: RAZ, AVRAHAM
: TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY

```

```

; TITLE OF INVENTION: OF METASTASIS IN A CELL SAMPLE
;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dykema Gossett
; STREET: STE 505 N. Woodward
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: U.S.
; ZIP: 48304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/562,311
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,225
; FILING DATE:
; APPLICATION NUMBER: US 07/681,242
; FILING DATE: 04-APR-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/294,249
; FILING DATE: 01-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: KELLY, ROBERT L.
; REGISTRATION NUMBER: 31,843
; REFERENCE/DOCKET NUMBER: 61,686-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-540-0849
; TELEFAX: 810-540-0763
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-562-311-4

Query Match          79.4%; Score 1078; DB 1; Length 264;
Best Local Similarity 78.2%; Pred. No. 4,3e-92;
Matches 208; Conservative 15; Mismatches 21; Indels 22; Gaps 8;

QY 1 MADNFSLDALSGSGNPNQGWPGWANGNPAGAGYGYPGASYPG-YFGQAPPGAYPGQADP 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 1 MADTFSLDALSGSGNPNQGWPGWANGNP-GAGGYPGAAYPGAYPGQADP 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 60 GAYHG--APGAYPGAGAPGVYPPGSPGAYPES-----GQSPAPGAY-----AT 102
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 60 GAYPGQAPPSAPPGPTAPGAYPG-PTAPGAYPGSTAPGAPFGQPGAPGAYPSPAPGSPAP 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 103 GPYGAGAPGLIYVYNLPLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFH-PREN 161
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 119 GPYGVAGPLIYVYDPLPGGLMRLITITMGVKNARIIVDFRGNDVAFHPRRN 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 162 ENNRRIYVNCIKDNNWGREEROSVFPFSSGKPEFKIQVLEPDHFVAVNDAAH-LOYNHR 220
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 179 ENNRRIYVNCIKDNNWGREEROSAPFPFSSGKPEFKIQVLEADHFVAVNDAAHLOYNHR 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 221 VKKLNISKLGISGDI.DLTSASYTMI 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 239 MKKLNREISQIGISGDI.TLTSANHAMI 264
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 7
US-08-647-960-5
; Sequence 5, Application US/08647960
; Patent No. 5908761
; GENERAL INFORMATION:
; APPLICANT: ZICK, Yehiel
; TITLE OF INVENTION: GALECTIN-8 AND GALECTIN-8-LIKE PROTEINS
```

```

; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,960
; FILING DATE: 30-MAY-1996
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 107880
; FILING DATE: 05-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: ZICK=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3527
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-647-960-5

Query Match          37.8%; Score 513; DB 2; Length 135;
Best Local Similarity 76.3%; Pred. No. 3,7e-40;
Matches 103; Conservative 11; Mismatches 19; Indels 2; Gaps 2;

QY 114 VPEYNLPLPGGVPRMLITLITGVKPNANRIALDFQGNDAVAFH-PRENNRRRIYVNCNT 172
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 1 VEYDMLPLPGGVPRMLITLITGVKPNANSETLINEKGNDAIASHENPRENNRRRIYVNCNT 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 173 KLDNNWGREEROSVFPFSSGKPEFKIQVLEPDHFVAVNDAAH-LOYNHVKKLNISKLG 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 61 KODNNWGREEROSAPFPFSSGKPEFKIQVLEADHFVAVNDVHLLQYNNHKKLNREISQIG 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 232 ISGDI.DLTSASYTMI 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 121 ITGDI.TLTSASHAMI 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 8
US-09-559-023-2
; Sequence 2, Application US/09559023
; Patent No. 6551796
; GENERAL INFORMATION:
; APPLICANT: Adamson, Ruth
; APPLICANT: Leal-Pinto, Edgar
; APPLICANT: Lipkowitz, Michael
; TITLE OF INVENTION: NUCLEIC ACID ENCODING URATE TRANSPORTER
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: 070165.0574
; CURRENT APPLICATION NUMBER: US/09/559,023
; CURRENT FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/221,898
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: US 60/099,752
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/070,215
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 8
```

SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 322
 TYPE: PRT
 ORGANISM: Rat
 US-09-559-023-2

Query Match 20.4%; Score 277.5; DB 4; Length 322;
 Best Local Similarity 32.9%; Pred. No. 6.9e-18;
 Matches 82; Conservative 29; Mismatches 67; Indels 71; Gaps 12;

QY FSLHDLGSG-----NPNQCPGAMGNQPAAGGYPGAYGQAPGAYPQQA 57
 DB 127 YHLVDTTISVSGCLHLSPINPQTQGF-----QPA-----HQA-----VAQT 162
 QY 58 PPGAHGAPG---AYGAPAPGVYGPSPGAYSSGQPSAPGAYATGPYAPAGAPGLIV 114
 DB 163 IITHVHSIFGQMLSTFGIIP-PMAYTPP-----AYT-----I 192
 QY 115 PYNLPDPGVVPRMLITLLGTVKPNANRIALDFQGNDAFHF--PRENNRRVIVCNTK 173
 DB 193 PFTSIPNGFYPSKINSIGVLLPDAKRFHINLRGGDIAFHINPRFE--KVVVRTQ 249
 QY 174 LDNNWGREERQ--SVFPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKLNLSKLT 230
 DB 250 INNWGPEERSLPRMPFRNGQSPFSVWILCEHGFKAVDGQHICEYHRLKNLPDINTL 309
 QY 231 GISGIDLT 239
 DB 310 EVAGDIQLT 318

RESULT 9

US-08-946-914-4
 Sequence 4, Application US/08946914
 Patent No. 6027916

GENERAL INFORMATION:
 APPLICANT: NI, Jlan

APPLICANT: Gentz, Reiner L.

APPLICANT: Ruben, Steven M.

TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Steiner, Kessler, Goldstein, & Fox P.L.L.C.

STREET: 1100 New York Ave., Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/946,914

FILING DATE: Herewith

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/028,093

FILING DATE: 09-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Steffe, Eric K.

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488, 0560001/EKS/SGM

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 311 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein
 US-08-946-914-4

Query Match 19.5%; Score 264; DB 3; Length 311;
 Best Local Similarity 36.5%; Pred. No. 1.2e-16;
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 APGAYPGAPA--PGVYGPSPGAYSSGQPSAPGAYATGPYAPAGAPGLIVPNLPLPG 122
 DB 159 APGGMSTPALPPMMYHP-----AYP-----MPPTTLLG 189
 QY 123 GVVPRMLITLLGTVKPNANRIALDFQGNDAFHF--PRENNRRVIVCNTKLDNNWGRE 181
 DB 190 GLYPSKSLILSGTLPDAQRFHINLCSGNHIAFHINPRFDE--AVVRNTQIDNSWGE 246
 QY 182 ERQ--SVFPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKLNLSKLTGISGIDL 238
 DB 247 ERSIPRKMFPVKGQSFVWILCEHGFKAVDGQHLEFYHRLKNLPDINTLEVGDIQL 306
 QY 239 T 239
 DB 307 T 307

RESULT 10

US-09-656-450-4
 Sequence 4, Application US/09656450
 Patent No. 6468768

GENERAL INFORMATION:
 APPLICANT: NI, Jlan

APPLICANT: Gentz, Reiner L.

APPLICANT: Ruben, Steven M.

TITLE OF INVENTION: Galectin 9 and 10SV polynucleotides

FILE REFERENCE: 1488, 0560003

CURRENT APPLICATION NUMBER: US/09/656,450

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: US 09/263,689

PRIOR FILING DATE: 1999-03-05

PRIOR APPLICATION NUMBER: US 08/946,914

PRIOR FILING DATE: 1997-10-09

PRIOR APPLICATION NUMBER: US 60/028,093

PRIOR FILING DATE: 1996-10-09

NUMBER OF SEQ ID NOS: 60

SOFTWARE: Patentin version 3.0

SEQ ID NO 4

LENGTH: 311

TYPE: PRT

ORGANISM: Homo sapiens

US-09-656-450-4

Query Match 19.5%; Score 264; DB 4; Length 311;
 Best Local Similarity 36.5%; Pred. No. 1.2e-16;
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 APGAYPGAPA--PGVYGPSPGAYSSGQPSAPGAYATGPYAPAGAPGLIVPNLPLPG 122
 DB 159 APGGMSTPALPPMMYHP-----AYP-----MPPTTLLG 189
 QY 123 GVVPRMLITLLGTVKPNANRIALDFQGNDAFHF--PRENNRRVIVCNTKLDNNWGRE 181
 DB 190 GLYPSKSLILSGTLPDAQRFHINLCSGNHIAFHINPRFDE--AVVRNTQIDNSWGE 246
 QY 182 ERQ--SVFPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKLNLSKLTGISGIDL 238
 DB 247 ERSIPRKMFPVKGQSFVWILCEHGFKAVDGQHLEFYHRLKNLPDINTLEVGDIQL 306
 QY 239 T 239
 DB 307 T 307

RESULT 11

US-08-788-584-5

```

; Sequence 5 Application US/08788584
; Patent No. 5837493
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goll, Surya K.
APPLICANT: Bandman, Olga
APPLICANT: Hawkins, Phillip R.
APPLICANT: Petichory, Joanne R.
TITLE OF INVENTION: NOVEL HUMAN GALECTINS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/788,584
FILING DATE: Filed Herewith
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0192 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 727176
US-08-788-584-E

Query Match 18.5%; Score 250.5; DB 2; Length 145;
Best Local Similarity 40.4%; Pred. No. 8e-16;
Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4,

QY 108 PAGELIVPYNILPLPGVVPKMLITLLGTVKKRNARIALDFQGNVAHFH--PRENNERR 166
      b 9 PYPMIAVFETTSIPGLYPSKSIVISGVALDAAKEFOINLRGGIATHLNRFEN-- 65
QY 167 VIVCNCTKDNDWGBEERO--SVFPESGKPEFKIQVLVEPDHFKXVANDAHL-QYNRYVK 223
      Db 66 AVAVNTQTINNNGPRPERSLPGSMPTSRGQRSVWLTCGHCKXVAVDQHICYSRLMN 125
QY 224 LNEISKLGISGDIDLT 239
      Db 126 LPDINTLEVAGDIQLT 141

RESULT 12
US-08-946-914-12
; Sequence 12, Application US/08946914
; Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
```

```

NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-946-914-12

Query Match 18.5%; Score 250.5; DB 3; Length 145;
Best Local Similarity 40.4%; Pred. No. 8e-16;
Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4

QY 108 PAGELIVPEYNLPFGGVVERMLTITIDGVKPNARIMLDFOGRNDVAHF--PREENNR 166
| | | | | : | | | | | : | | | | | : | | | | |
DB 9 PYPMLAPPFPSISINGYPSKSIYISGVVLSDAKRFQINRCGDIAFHLNFRDEN--- 65

QY 167 VIVONTLDDNNWGEHQ--SVFPEESGKPEKIVLVEPDHFQAVYNDAAH-QYNHRYKK 223
| | | | | : | | | | | : | | | | | : | | | | |
DB 66 AVANNNTINNSWGEERSLPSGMPFSRGQRFVWLCEGHCFKVAVDQGHICEYSHRLMN 125

QY 224 LNEISKLGISGIDILT 239
| | | | | : | | | | |
DB 126 LPDINTLEVAGDIQILT 141

RESULT 13
US-09-656-450-12
; Sequence 12, Application US/09656450
; Patent No. 6468768
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruden, Steven M.
; TITLE OF INVENTION: Galleclin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560003
; CURRENT APPLICATION NUMBER: US/09/656,450
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1993-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; PRIOR FILING DATE: 1996-10-09
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patentin version 3.0

```


;; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE

;; FILE REFERENCE: 210121.471C2

;; CURRENT APPLICATION NUMBER: US/09/401.064

;; CURRENT FILING DATE: 1999-09-22

;; NUMBER OF SEQ ID NOS: 371

;; SOFTWARE: FastSeq for Windows Version 3.0

;; SEQ ID NO: 199

;; LENGTH: 168

;; TYPE: PRT

;; ORGANISM: Homo sapien

US-09-401-064-199

Query Match 18.2%; Score 246.5; DB 4; Length 168;

Best Local Similarity 37.3%; Pred. No. 2.2e-15;

Matches 59; Conservative 33; Mismatches 51; Indels 15; Gaps 8;

QY 92 GQPSAPGAVATGPGAGPLIVPNLPLPGGVPRMLITLIGTV-KENANRIALDFQ-- 148

DB 5 GEMATSGQA--PYLSPA---VPFSGTIGGGLQDGLQITVNGTVLSSSGTRFAVNFQTG 58

QY 149 -RGNDVAFHF-PRFNNRNRIVVNTKLDNMWGREEROSVFPFESGKPFKIQVLPDPHF 206

DB 59 FSGNDIAHFHFPRFDG--YVVCNTRQNGSWGPEERKTHMPFQKMPFDLCFIVQSSDF 116

QY 207 KVAVND-AHLOYNHRVKKLNEISKLGISGDIPLTSASY 243

DB 117 KVMVNGILFVQYFHFV-PPHRVDTISVNGSVQLSTISF 153

RESULT 17

US-08-788-584-1

Sequence 1, Application US/08788584

Patent No. 5837493

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Goli, Surya K.

APPLICANT: Bandman, Olga

APPLICANT: Hawkins, Phillip R.

APPLICANT: Petichory, Joanne R.

TITLE OF INVENTION: NOVEL HUMAN GALECTINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/788,584

FILING DATE: Filed Herewith

CLASSIFICATION: 436

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0192 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-845-4166

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-788-584-1

Query Match 17.9%; Score 243.5; DB 2; Length 145;

Best Local Similarity 40.0%; Pred. No. 3.5e-15;

Matches 56; Conservative 23; Mismatches 50; Indels 11; Gaps 5;

QY 104 PYGAPAGPLIVPNLPLPGGVPRMLITLIGTV-KENANRIALDFQ--RGNDVAFHF-PRNE 162

DB 9 PYLSPX---VPFSGTIGGGLYPSKSTILSGTVLPSAQRFHINLCGSHIGFHLNPRDE 64

QY 163 NNRRVIVVNTKLDNMWGREERQ--SVFPFESGKPFKIQVLPDPHFVAVNDALH-QYNH 219

DB 65 N---AVVANMQIDNMWGREERSLPRKMPFVVGQSFVWLCEAHGLKVAVDGQHLEFYH 121

QY 220 RVKLNLEISKLGISGDIPLT 239

DB 122 RLRLPLTINRLVGGDIQLT 141

RESULT 18

US-09-559-023-4

Sequence 4, Application US/09559023

Patent No. 6551796

GENERAL INFORMATION:

APPLICANT: Abramson, Ruth

APPLICANT: Leal-Pinto, Edgar

APPLICANT: Lipkowitz, Michael

TITLE OF INVENTION: NUCLEIC ACID ENCODING URATE TRANSPORTER

FILE REFERENCE: 070165.0574

CURRENT APPLICATION NUMBER: US/09/559,023

CURRENT FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: US 09/221,898

PRIOR FILING DATE: 1998-12-28

PRIOR APPLICATION NUMBER: US 60/099,752

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: US 60/070,215

PRIOR FILING DATE: 1997-12-31

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 301

TYPE: PRT

ORGANISM: Human

US-09-559-023-4

Query Match 17.9%; Score 243.5; DB 4; Length 301;

Best Local Similarity 38.4%; Pred. No. 8.9e-15;

Matches 56; Conservative 30; Mismatches 47; Indels 13; Gaps 7;

QY 104 PYGAPAGPLIVPNLPLPGGVPRMLITLIGTV-KENANRIALDFQ--RGNDVAFHF-P 158

DB 9 PYLSPA---VPFSGTIGGGLQDGLQITVNGTVLSSSGTRFAVNFQTGSSNDIAHFNP 64

QY 159 RENNNRNVIVVNTKLDNMWGREEROSVFPFESGKPFKIQVLPDPHFVAVND-AHLOY 217

DB 65 REDDGG--YVVCNTRQNGSWGPEERKTHMPFQKMPFDLCFIVQSSDFVAVNGILFVQY 122

QY 218 NRVYKLNLEISKLGISGDIPLTSASY 243

DB 123 FHRV-PPHRVDTISVNGSVQLSTISF 147

RESULT 19

US-08-946-914-11

Sequence 11, Application US/08946914

Patent No. 6027916

GENERAL INFORMATION:

APPLICANT: Ni, Jian

APPLICANT: Gentz, Reiner L.

APPLICANT: Ruben, Steven M.

TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
 STREET: 1100 New York Ave., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/946,914
 FILING DATE: Herewith
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/028,093
 FILING DATE: 09-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Steffe, Eric K.
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-2600
 TELEFAX: 202-371-2540
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 324 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-946-914-11

Query Match 17.4%; Score 236.5; DB 3; Length 324;
 Best Local Similarity 35.2%; Pred. No. 4.3e-14;
 Matches 68; Conservative 25; Mismatches 63; Indels 37; Gaps 10;

QY 71 GAPPAGVPPSPGPGAPSSGQSPAPGAVATGPYAP-----AGPLI-----VPYNL 118
 DB 152 GPPASQYPTMTIP-AVPSAG-----YNPQNMSLPYMAGPPIFNPPVYVG 198
 QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQRGN--DVAHF--PRENNRRVIVCNTKLD 175
 DB 199 TLGGGLTARRITIIKGYVLPTAKNLIINFKVGSTGDIAFHNNPRIGD----CVARNSTYN 254
 QY 176 NNMGREERQSVF-PFESGKPEFKIQLVLPDHFKAIVNDAHL-QYNHRVKKLINEISKGIS 233
 DB 255 GSWSEBRKIPYNPFAGQGFDDLSIRCGTDRFKVFANGQHLDFSHRFQAFQVDMLEIK 314
 QY 234 GDIDLTSASYTMI 246
 DB 315 GDITL---SYVQI 324

RESULT 20
 US-09-656-450-11
 Sequence 11, Application US/09656450
 Patent No. 6468768
 GENERAL INFORMATION:
 APPLICANT: Ni, Jian
 APPLICANT: Gentz, Reiner L.
 APPLICANT: Ruben, Steven M.
 TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
 FILE REFERENCE: 1488.0560003
 CURRENT APPLICATION NUMBER: US/09/656,450
 FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: US 09/263,689
 PRIOR FILING DATE: 1999-03-05
 PRIOR APPLICATION NUMBER: US 08/946,914
 PRIOR FILING DATE: 1997-10-09
 PRIOR APPLICATION NUMBER: US 60/028,093

PRIOR FILING DATE: 1996-10-09
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: Patent version 3.0
 SEQ ID NO 11
 LENGTH: 324
 TYPE: PRT
 ORGANISM: Rat
 US-09-656-450-11

Query Match 17.4%; Score 236.5; DB 4; Length 324;
 Best Local Similarity 35.2%; Pred. No. 4.3e-14;
 Matches 68; Conservative 25; Mismatches 63; Indels 37; Gaps 10;

QY 71 GAPPAGVPPSPGPGAPSSGQSPAPGAVATGPYAP-----AGPLI-----VPYNL 118
 DB 152 GPPASQYPTMTIP-AVPSAG-----YNPQNMSLPYMAGPPIFNPPVYVG 198
 QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQRGN--DVAHF--PRENNRRVIVCNTKLD 175
 DB 199 TLGGGLTARRITIIKGYVLPTAKNLIINFKVGSTGDIAFHNNPRIGD----CVARNSTYN 254
 QY 176 NNMGREERQSVF-PFESGKPEFKIQLVLPDHFKAIVNDAHL-QYNHRVKKLINEISKGIS 233
 DB 255 GSWSEBRKIPYNPFAGQGFDDLSIRCGTDRFKVFANGQHLDFSHRFQAFQVDMLEIK 314
 QY 234 GDIDLTSASYTMI 246
 DB 315 GDITL---SYVQI 324

RESULT 21
 US-08-946-914-8
 Sequence 8, Application US/08946914
 Patent No. 6027916
 GENERAL INFORMATION:
 APPLICANT: Ni, Jian
 APPLICANT: Gentz, Reiner L.
 APPLICANT: Ruben, Steven M.
 TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
 NUMBER OF SEQUENCES: 60
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
 STREET: 1100 New York Ave., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/946,914
 FILING DATE: Herewith
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/028,093
 FILING DATE: 09-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Steffe, Eric K.
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-2600
 TELEFAX: 202-371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 200 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-946-914-8

Query Match	16.5%;	Score 223.5;	DB 3;	Length 200;
Best Local Similarity	39.3%;	Pred. No. 3,8e-13;		
Matches 55;	Conservative 23;	Mismatches 51;	Indels 11;	Gaps 5

QY 113 IVPENLPLPGGVNVRMLTLITIGVYENANRRLALDFQGN-----DVAEHF-ERFENENR 165
 Db 16 VIPPVCTIPDGLPCTLLVIRGHVPSDDADRFDVLDQNGSSMMKPRADVAFHFMRPRKRGAG- 74
 QY 166 RVIACNTKLDNNMGHEERQSVFPPESGKPKFKIQLVVEPDHFKVAVNDAAH-LQYNHRVKKL 224
 Db 75 -CIYCNTILNKKRGRELLTYDTPFRERSFELVIMVLDKKPVAVANCKRTLLYGRIGD- 132
 QY 225 NEISKLGISGDIIDLTSASYT 244
 Db 133 EKIDPLTIGYKVNTHSIGFS 152

RESULT 22
US-09-656-450-8
; Sequence 8, Application US/09656450

```

1  APPLICANT: Ni, Jian
2  APPLICANT: Gentz, Reiner L.
3  APPLICANT: Ruben, Steven M.
4  TITLE OF INVENTION: Galactin 9 and 10SV Polynucleotides
5  FILE REFERENCE: 1488,0560003
6  CURRENT APPLICATION NUMBER: US/09/656,450
7  CURRENT FILING DATE: 2000-09-06
8  PRIOR APPLICATION NUMBER: US 09/263,689
9  PRIOR FILING DATE: 1999-03-05
10 PRIOR APPLICATION NUMBER: US 08/946,914
11 PRIOR FILING DATE: 1997-10-09
12 PRIOR APPLICATION NUMBER: US 60/028,093
13 PRIOR FILING DATE: 1996-10-09
14 NUMBER OF SEQ ID NOS: 60
15 SOFTWARE: PatentIn version 3.0
16 SEQ ID NO: 8
17 LENGTH: 200
18 TYPE: PRT
19 ORGANISM: Homo sapiens
20 US-09-656-450-8

```

Query Match	16.5%;	Score 223.5;	DB 4;	Length 200;
Best Local Similarity	39.3%;	Pred. No. 3.8e-13;		
Matches	55;	Conservative	23;	Mismatches 51;
				Indels 11;
				Gaps 5

```
QY 113 IVEYNLPFEGGVVPRMLITILGVKNPNAIADFORGN-----DVAFHF-PRFNNNR 165
Db 16 VLPFVGITDQDLPGLTIVIRGHPBDADAFQVDLNGSSMKRADAFFHNRRFKAG- 74
QY 166 RYIVCNCTKLDNMNGREERQSVPFESGKPFKIQVIVEPHEKVAVNDAH-LQYNHRVYKL 224
Db 75 -CIVCNTLINEKGRREIITYDPPFKRKSEFIIVMLKDKFQVAVNGKHTLLIGHRIQP- 134
QY 225 NEISKLGISGIDILNTASYT 244
Db 133 EKIDITLGIGKNNIHSIGFS 152
```

RESULT 23
 US-09-131-648-5
 ; Sequence 5, Application US/09131648
 ; Patent No. 6168920
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Guedler, Karl J.
 ; APPLICANT: Patterson, Chandra
 ; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
 ; FILE REFERENCE: PR-0576 US

CURRENT APPLICATION NUMBER: US/09/131,648
CURRENT FILING DATE: 1998-08-10

```

SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: g1932712
US-09-131-648-5

```

Query Match	16.5%;	Score 223.5;	DB 3;	Length 316;
Best Local Similarity	39.3%;	Pred. No. 6.7e-13;		
Matches	55;	Conservative	23;	Mismatches 51;
				Indels 11;
				Gaps 5;

QY	113	IYDYNLPLPGSVAPRMLLITLIGTVKPNRALTDFQGN-----DVAFFH-PRFENNNR	165
		: : : : :	
Db	15	VIFVFGTIPDQLPGTLVIRGHVPSDDRRQVLDLQNGSVKPEADVAHFHNRPRFKAG-	73
		: : : : :	
QY	166	RVVNCCTKLDNNWGREGQSVPRFESGKPRFIQVLVEPDHFKVAVNDAN-LQYNHRVYKL	224
		: : : : : : : : : :	
Db	74	-CTVNCNTLNEKMGREBITYDTPFRKESFELVIMVLTCKRQVAVNGKHTLLVGRIGP-	121
		: : : : : : : : : :	
QY	225	NEISKLGISGIDILTSASYT	244
		: : : : :	
Db	132	EKIDTLGITYGVNIHSIGFS	151
		: : : : :	

RESULT 24
US-08-946-914-6
; Sequence 6, Application US/08946914
; Patent NO. 6027916

APPLICANT: NI, Jian
 APPLICANT: Gentz, Reiner L.
 APPLICANT: Rubert, Steven M.
 TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
 NUMBER OF SEQUENCES: 60
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Steneme, Kessler, Goldstein, & Fox P.L.L.C.
 STREET: 1100 New York Ave., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/946,914
 FILING DATE: Herewaltch
 CLASSIFICATION: 530
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 60/028,093
 FILING DATE: 09-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Steffe, Eric K.
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-2600
 TELEFAX: 202-371-2540
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 317 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match	16.5%;	Score 223.5;	DB 3;	Length 317;
Best Local Similarity	39.3%;	Pred. No. 6.7e-13;		
Matches 55; Conservative	23;	Mismatches 51;	Indels 11;	Gaps 5

```

QY      113  IVVYNPLPEGVVVERMLITILGVKNARMLDPRGN-----DVAFEH-PPFENNR 165
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      16  VIFPVGTIEQDLPGLTILVIRGVESDADRFOVDLQNSQSVKPRADVAFHNPFRFKAG- 74

QY      166  RVIVCNCTKLNNNGREERQSVFPFESGPKFQVLVEBDHKVAVNDAH-LQNNHRYKL 224
      |||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      75  -CIVCNITLINEKMGREIITYDPFRFKRSFELIVLWLDKQVAVNGKHTLTYGHRIGP- 132

QY      225  NEISKLGISGIDILDTASYT 244
      : ||| :: :: :: ::
Db      133  EKIDTGLGYKVNHIHSIGFS 152

```

```

RESULT 25
US-09-656-450-6
; Sequence 6, Application US/09656450
; Patent No. 6468768
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galactin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560003
; CURRENT APPLICATION NUMBER: US/09/656,450
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; PRIOR FILING DATE: 1996-10-09
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-656-450-6

```

	Query Match	16.5%	Score 223.5	DB 4	Length 317
	Best Local Similarity	39.3%	Pred No. 6.7e-13		
	Matches	55	Conservative	23	Mismatches 51; Indels 11; Gaps 5
QY	113	IVPNLPPGGVPEMLTILITGVKPNANRIALDPFGN-----DVAHF-F-DRFENNRR	165		
		::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	16	VAPFGVTPDDQDPETLLVIRGHWPSDDRRQVDLQNGSSVKPRADVAFFHNFRFRAG-	74		
QY	166	RIYVNTKLDNNWGEERQSVFPESGKEFIQVLVPEDFHKAANDAH-LQYNHRVKLL	224		
		:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	75	-CIQCNNTLINEKMGEEITYDTPFERKESFEIVIMVLKDFEQYAVNGKHTLLVGHRIGP-	132		
QY	225	NEISLIGSIDIDLTSASYT	244		
		::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	133	EKIDTPLGIVGKVNIIHSIGFS	152		

RESULT 26
US-08-728-521-3
; Sequence 3, Application US/08728521
; Patent No. 5869289
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Philip R.
; TITLE OF INVENTION: HUMAN GALECTIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Pottery Drive
; CITY: Palo Alto

```

? STATE: CA
? COUNTRY: US
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq Version 1.5
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/728,521
? FILING DATE: Filed Herewith
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0137 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-8166
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 316 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? IMMEDIATE SOURCE:
? LIBRARY: GenBank
? CLONE: 717032

```

[illegible]

RESULT 27
 US-08-647-960-2
 ; Sequence 2, Application US/08647960
 ; Patent No. 5908761
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: ZICK, Yehiel
 ; TITLE OF INVENTION: GALECTIN-8 AND GALECTIN-8-LIKE PROTEINS
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ;
 ; ZIP: 20004
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/647,960
 ; FILING DATE: 30-MAY-1996

CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107880
FILING DATE: 05-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: ZICK=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3527
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-647-960-2

Query Match 16.1%; Score 218.5; DB 2; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPYNLPDGVVPRMLITILGTVKPNRRIALDPQRGN-----DVAFHF-PRFENNRR 166
DB 16 IPYVSTIEQLKPGSLIYIRGHVPKDSERFQVDFQHGSLSKPRADVAHFHNPFRKRSN-- 73
QY 167 VIVCNITKLDNNWGEROSVFPFESGKPEKIQVLYEPDHEKAVANDAH-LQYHRYVKLN 225
DB 74 CIVCNLTINERKMGWEIITHDMFPRKESFEIIVWLKNKFVAVANGKHILLIYAHRIINP-E 132
QY 226 EISKLGISGDIDILTSASY 243
DB 133 KIDTLGIPIGKAVIHISIGF 150

RESULT 28
US-08-946-914-15
Sequence 15, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-15

Query Match 16.1%; Score 218.5; DB 3; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPYNLPDGVVPRMLITILGTVKPNRRIALDPQRGN-----DVAFHF-PRFENNRR 166
DB 16 IPYVSTIEQLKPGSLIYIRGHVPKDSERFQVDFQHGSLSKPRADVAHFHNPFRKRSN-- 73
QY 167 VIVCNITKLDNNWGEROSVFPFESGKPEKIQVLYEPDHEKAVANDAH-LQYHRYVKLN 225
DB 74 CIVCNLTINERKMGWEIITHDMFPRKESFEIIVWLKNKFVAVANGKHILLIYAHRIINP-E 132
QY 226 EISKLGISGDIDILTSASY 243
DB 133 KIDTLGIPIGKAVIHISIGF 150

RESULT 29
US-08-946-914-17
Sequence 17, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-946-914-17

Query Match 16.1%; Score 218.5; DB 3; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

```

Qy      114 VPNALPPGGVVPRLTTITIGTVKPNANRIALDFGRN-----DVAFHF-PRFENNRR 166
       :|||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      16 IPVSTITEQLTKSGSLIVIRGHVPSERFQVDFQHGNLSLKPRADVAHFHNPFRKSN-- 73
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      167 VIVCNTKLDDNWGSEEROSVFPEESGKPFIQVLVEPDHFKAVDNAH-LQYHRVKLN 229
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      74 CIVCNTLLNEKMGMEETIHDMPFRKEKSFELIVIMVLKNKHVAVNGHILLTAHRINP-E 132
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      226 EISKIGISGDIDLTASAY 243
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      133 KIDTTLGIFFKVHTHSIGF 150
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 30
US-09-212-146-3
; Sequence 3, Application US/09212146
; Patent No. 628133
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: HUMAN GALECTIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/212,146
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/728,521
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0137 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 717032
; US-09-212-146-3

Query Match          16.1%; Score 218.5; DB 3; Length 316;
Best Local Similarity 39.9%; Pred.No.1,9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5

Qy      114 VPYNLPLRGVVRPLITILITGVKPNANRIALDFGRN-----DVAFHF-PRFENNRR 166
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      16 IPVSTITEQLTKSGSLIVIRGHVPSERFQVDFQHGNLSLKPRADVAHFHNPFRKSN-- 73
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      167 VIVCNTKLDDNWGSEEROSVFPEESGKPFIQVLVEPDHFKAVDNAH-LQYHRVKLN 225
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      74 CIVCNTLLNEKMGMEETIHDMPFRKEKSFELIVIMVLKNKHVAVNGHILLTAHRINP-E 132
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      226 EISKIGISGDIDLTASAY 243
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      133 KIDTTLGIFFKVHTHSIGF 150
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||


```

```

      : ||| |::||:
Db      133 KIDTLGIFGKVNHTSIGF 150

      RESULT 31
US-09-656-450-15
; Sequence 15, Application US/09656450
; Patent No. 6468768
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560003
; CURRENT APPLICATION NUMBER: US/09/656,450
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; PRIOR FILING DATE: 1996-10-09
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Rat
US-09-656-450-15

Query Match          16.1%; Score 218.5; DB 4; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 34; Indels 11; Gaps 5

QY      114 VPVNLPLDGGVVPRLITILCTGVKNANRIALDFQPGN-----DVAFHF--PRENNNR 166
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      16 IPYSTINEQLKPSGLIYIRGHVPRDSERFQVDFFHSGSLKRPADVAPHFPRFRSN-- 73
QY      167 VIVCNCKLDNNWNGEEROSVPPFSKGFEFKIQVLVEPHEFYAVANDAH-LQYNRVKKLN 225
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      74 CIVCNTLTNEKGMGEELTHDMPEFRKEKSFEIVIMWLKNKFVAVANGKIILLYHRIND-E 132
QY      226 EISKLGISGDIDLTSASY 243
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      133 KIDTLGIFGKVNHTSIGF 150

RESULT 32
US-09-656-450-17
; Sequence 17, Application US/09656450
; Patent No. 6468768
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560003
; CURRENT APPLICATION NUMBER: US/09/656,450
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; PRIOR FILING DATE: 1996-10-09
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Rat
US-09-656-450-17
```

Query Match 16.1%; Score 218.5; DB 4; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPVNLPLPGVPPMLTITLIGTVKPNANRIALDFQVFNHPRFNNRR 166
DB 16 IPVSTTEQLKPSGLVIRGHVPEKDSRFQVDFHNSLKPRAVAFHPRKRSN-- 73
QY 167 VIVCNKLDNNWGREEROSVPEPESGKPFKIQVLEPHFVAVNDAAH-LQYNERVKLN 225
DB 74 CIVCNLTLENKGMWEITHDMPEKESFEIIVMLKKNFVAVNGKILLVYHRINP-E 132
QY 226 EISKIGISGDIIDLTASY 243
DB 133 KIDTLGIFGKVNIHSGF 150

RESULT 33

US-08-469-667-16
Sequence 16, Application US/08469667
Patent No. 5733748
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gillfillan, Cecchi,
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,667
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-667-16

Query Match 15.9%; Score 215.5; DB 1; Length 323;
Best Local Similarity 34.4%; Pred. No. 3.8e-12;
Matches 52; Conservative 28; Mismatches 62; Indels 9; Gaps 5;

QY 99 AYATGPGAGPLIVPNLPLPGVPPMLTITLIGTVKPNANRIALDF----QRGNDVA 154
DB 2 AYVPAAGYQPTNPPLIPYQPIPGGLNVGMSVYIQVASEHMKRFVNFVVGDDPSDVA 61
QY 155 FHF-PRENNRRVIVCNKLDNNWGREEROSVPEPESGKPFKIQVLEPHFVAVN-D 212
DB 62 FHNPRFDGMDK--VFNTLQGGKWSERKSMPEKGAFAELVPLVLAHRYKVVNGN 119
QY 213 AHIQYHRYKKNISKLIGISGDIIDLTASY 243
DB 120 PFEYEGHRL-PLQWVTHLQVDDGLQLOSINF 149

RESULT 34

US-08-946-914-2
Sequence 2, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steine, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-2

Query Match 15.9%; Score 215.5; DB 3; Length 323;
Best Local Similarity 34.4%; Pred. No. 3.8e-12;
Matches 52; Conservative 28; Mismatches 62; Indels 9; Gaps 5;

QY 99 AYATGPGAGPLIVPNLPLPGVPPMLTITLIGTVKPNANRIALDF----QRGNDVA 154
DB 2 AYVPAAGYQPTNPPLIPYQPIPGGLNVGMSVYIQVASEHMKRFVNFVVGDDPSDVA 61
QY 155 FHF-PRENNRRVIVCNKLDNNWGREEROSVPEPESGKPFKIQVLEPHFVAVN-D 212
DB 62 FHNPRFDGMDK--VFNTLQGGKWSERKSMPEKGAFAELVPLVLAHRYKVVNGN 119
QY 213 AHIQYHRYKKNISKLIGISGDIIDLTASY 243
DB 120 PFEYEGHRL-PLQWVTHLQVDDGLQLOSINF 149

RESULT 35

US-09-224-110-16
Sequence 16, Application US/09224110
Patent No. 6337195
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:

ADDRESSER: Carella, Byrne, Bain, Gillfillan, Cecchi,
ADDRESSER: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/224,110
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/469,667
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-224-110-16

Query Match 15.9%; Score 215.5; DB 4; Length 323;
Best Local Similarity 34.4%; Pred. No. 3,8e-12;
Matches 52; Conservative 28; Mismatches 62; Indels 9; Gaps 5;

QY 99 AYATGPGAPAGPLIVPNLPLPGGVVERMLITLITGVKPNANRIALDF---ORGNDVA 154
Db 2 AYVAPAGYQPTYNPTLPYQPIPGGLNVGMSVYIQGVASEHMKRFVFNFFVVGQDPSGSDVA 61

QY 155 FHF-PRFENNRRIVYCNCTKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKVAVN-D 212
Db 62 FHFNRPRDGMDC--VFENITLOGGKWSERKSRMPFKGAFAELVFIvlaEHYKVVVNGN 119

QY 213 AHLQYNRVKKLNEISKLGISGDIIDLSASY 243
Db 120 PFVEYGHRL-PLQWVTHLQVDDIQLQSIINF 149

RESULT 36
US-09-656-450-2
Sequence 2, Application US/09656450
Patent No. 6468768
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 9 and 10SV polynucleotides
FILE REFERENCE: 1488,056003
CURRENT APPLICATION NUMBER: US/09/656,450
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/263,689
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: US 08/946,914
PRIOR FILING DATE: 1997-10-09
PRIOR APPLICATION NUMBER: US 60/028,093
PRIOR FILING DATE: 1996-10-09
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 323

TYPE: PRT
ORGANISM: Homo sapiens
US-09-656-450-2

Query Match 15.9%; Score 215.5; DB 4; Length 323;
Best Local Similarity 34.4%; Pred. No. 3,8e-12;
Matches 52; Conservative 28; Mismatches 62; Indels 9; Gaps 5;

QY 99 AYATGPGAPAGPLIVPNLPLPGGVVERMLITLITGVKPNANRIALDF---ORGNDVA 154
Db 2 AYVAPAGYQPTYNPTLPYQPIPGGLNVGMSVYIQGVASEHMKRFVFNFFVVGQDPSGSDVA 61

QY 155 FHF-PRFENNRRIVYCNCTKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKVAVN-D 212
Db 62 FHFNRPRDGMDC--VFENITLOGGKWSERKSRMPFKGAFAELVFIvlaEHYKVVVNGN 119

QY 213 AHLQYNRVKKLNEISKLGISGDIIDLSASY 243
Db 120 PFVEYGHRL-PLQWVTHLQVDDIQLQSIINF 149

RESULT 37
PCT-US95-07289-16
Sequence 16, Application PC/TUS9507289
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Carella, Byrne, Bain, Gillfillan, Cecchi,
ADDRESSER: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07289
FILING DATE: 06-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-265
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-07289-16

Query Match 15.9%; Score 215.5; DB 5; Length 323;
Best Local Similarity 34.4%; Pred. No. 3,8e-12;
Matches 52; Conservative 28; Mismatches 62; Indels 9; Gaps 5;

QY 99 AYATGPGAPAGPLIVPNLPLPGGVVERMLITLITGVKPNANRIALDF---ORGNDVA 154
Db 2 AYVAPAGYQPTYNPTLPYQPIPGGLNVGMSVYIQGVASEHMKRFVFNFFVVGQDPSGSDVA 61

QY 155 FHF-PRFENNRRIVYCNCTKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKVAVN-D 212
Db 62 FHFNRPRDGMDC--VFENITLOGGKWSERKSRMPFKGAFAELVFIvlaEHYKVVVNGN 119

```
QY      213 AHLQYNHRVKKLTNEISKGLSGDIDLTASY 2433  
          :|::|::|::|::|::|:  
Db      120 PFYEYGHRL-PLQWTHLQVDGDILQLQSINF 1493
```

RESULT 38
US-08-946-914-13
; Sequence 13, Application US/08946914

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

RESULT 39
US-09-154-750A-79
; Sequence 79, Application US/09154750A
; Patent No. 6432640
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert

```

1  APPLICANT: Kluetzler, Kenneth
2  APPLICANT: Polyak, Kornelia
3  TITLE OF INVENTION: p53-Induced Apoptosis
4  FILE REFERENCE: 1107/75357
5  CURRENT APPLICATION NUMBER: US/09/154,750A
6  CURRENT FILING DATE: 1998-09-17
7  PRIOR APPLICATION NUMBER: 60/059,153
8  PRIOR FILING DATE: 1997-09-17
9  PRIOR APPLICATION NUMBER: 60/079817
10  PRIOR FILING DATE: 1998-03-30
11  NUMBER OF SEQ ID NOS: 93
12  SOFTWARE: FastSeq for Windows Version 3.0
13  SEQ ID NO: 79
14  LENGTH: 136
15  TYPE: PRT
16  ORGANISM: Homo sapiens
17  US-09-154-750A-79

```

Query Match	15.5%;	Score 211;	DB 4;	Length 136;
Best Local Similarity	35.3%;	Pred. No. 3.3e-12;		
Matches 47;	Conservative 28;	Mismatches 48;	Indels 10;	Gaps 5

```

RESULT 40
US-09-656-450-13
; Sequence 13, Application US/09656450
; Patent No. 6468768
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruban, Steven M.
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1489.0560003
; CURRENT APPLICATION NUMBER: US/09/656,450
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; PRIOR FILING DATE: 1996-10-09
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 136
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-656-450-13

```

Query Match 15.5%; Score 211; DB 4; Length 136;
Best Local Similarity 35.3%; Pred. No. 3.3e-12;
Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;

[illegible]

QY 228 SKLGISGIDILTS 240
DB 120 RLVWGVGDVQLDS 132

RESULT 41

US-09-557-170A-4
; Sequence 4, Application US/09557170A
; Patent No. 6605699
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Galectin 11
; FILE REFERENCE: PR354P2
; CURRENT APPLICATION NUMBER: US/09/557,170A
; CURRENT FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/109,864
; PRIOR FILING DATE: 1998-06-06
; PRIOR APPLICATION NUMBER: 09/010,146
; PRIOR FILING DATE: 1998-01-21
; PRIOR APPLICATION NUMBER: 60/034,205
; PRIOR FILING DATE: 1997-01-21
; PRIOR APPLICATION NUMBER: 60/034,204
; PRIOR FILING DATE: 1997-01-21
; PRIOR APPLICATION NUMBER: 60/169,932
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/130,390
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-557-170A-4

Query Match 15.3%; Score 207; DB 4; Length 318;
Best Local Similarity 38.3%; Pred. No. 2.3e-11;

Matches 54; Conservative 23; Mismatches 52; Indels 12; Gaps 6;

QY 113 IVPYNLPPLGGVPRMLTITIGVFNANRIALDFORGN-----DVAHF-PEFENNR 165
DB 16 VIPFVGTIPDQLDPTGLIVIRGHPSDADRFOVDLQNGSSMKPRADVAFHNPFFKAG- 74
QY 166 RVTYCNKTLDNMNGREERQSVFPE-SGKPKIOLVPEDPHFKAANDAH-LQNHRYKK 223
DB 75 -CIVCNLTINKMGREBELTYDTPFOKEKSFELIYIMLKAFQVAVNGKHTLLYGHRI GP 133
QY 224 LNEISKLGISGIDILTSASYT 244
DB 134 -EKIDTIGIYKVAIHIGFS 153

RESULT 42

US-09-219-849-3
; Sequence 3, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOORBOEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentln Ver. 2.1

; SEQ ID NO 3
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
US-09-219-849-3

Query Match 14.8%; Score 200.5; DB 3; Length 504;
Best Local Similarity 49.2%; Pred. No. 1.6e-10;

Matches 62; Conservative 1; Mismatches 50; Indels 13; Gaps 9;

QY 8 HDALSGSGNPNPGWPGAMQNPAGAGVPG-ASYPGVPGQA-PRGAVPGQA-PPGA--Y 62
DB 12 HGPAGPKAPGAPGAPGAPG--PACPPGAPGAPGAPGAPGAPGAPGAPGAPGAP 68
QY 63 HGAPGAYGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 120
DB 69 AGPPGA-PGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 125
QY 121 PGVVP 126
DB 126 PGAPGP 131

RESULT 43

US-08-642-255-52
; Sequence 52, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Berttram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A5556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-642-255-52

Query Match 14.8%; Score 200.5; DB 1; Length 561;
Best Local Similarity 49.2%; Pred. No. 1.9e-10;
Matches 62; Conservative 1; Mismatches 50; Indels 13; Gaps 9;

ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 777 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-53

Query Match
Best Local Similarity 49.2%; Score 199.5; DB 1; Length 777;
Matches 62; Conservative 1; Mismatches 50; Indels 13; Gaps 9;

QY 8 HDALSGSNPNQWPGAMGNQPAAGAGYPG-ASYPGYPGA-PGAYPGA-PPGA--Y 62
DB 45 HGPAGPKAPGPAGPGAPG-PAQPPGAPGPAGPPGAPGAPGAPGAPGAPGAPGAP 101
QY 63 HGAPAYGAPAPGVPYPPGPPSGGAYPSSGQSPAPGAVATGPGYGA--PAGPLIVPNLPL 120
DB 102 AGPPGA-PGPAGPPGAPGPAGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 158
QY 121 PGGVVP 126
DB 159 PGAGP 164

RESULT 47
US-08-642-255-51
Sequence 51, Application US/08642255
Patent No. 5773249

GENERAL INFORMATION:
APPLICANT: CAPPELO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-642-255-51

Query Match
Best Local Similarity 14.6%; Score 197.5; DB 1; Length 234;
Matches 61; Conservative 1; Mismatches 44; Indels 13; Gaps 9;

QY 15 GNPNGWPGAMGNQPAAGAGYPG-ASYPGYPGA-PGAYPGA-PPGA--YHGAPEAY 69
DB 1 GAPGAPGPPGAPG-PAQPPGAPGPAGPPGAPGAPGAPGAPGAPGAPGAPGAPGAP 56
QY 70 PGAPAGVYPPGPPSGGAYPSSGQSPAPGAVATGPGYGA--PAGPLIVPNLPLPGVVP 126
DB 57 PGAPGPPAPGPAGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 113

RESULT 48
US-08-526-136-13
Sequence 13, Application US/08526136
Patent No. 6107089

GENERAL INFORMATION:
APPLICANT: Towle, Christine A. et al.
TITLE OF INVENTION: ANNEXIN XI
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,136
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,036
FILING DATE:
APPLICATION NUMBER: 07/837,775
FILING DATE: February 13, 1992
APPLICATION NUMBER: 07/764,465
FILING DATE: September 23, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/099001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 466
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-526-136-13

Query Match
Best Local Similarity 14.5%; Score 196.5; DB 3; Length 466;
Matches 57; Conservative 6; Mismatches 45; Indels 39; Gaps 11;

QY 17 PNPQWPGAMGNQPAAGAGYPGAYPGAPGAYPGA-PGAYPGA-PGAYPGA-PGAYPGA 76
DB 33 PYPSPG-----PMGGATPYPVPSSGYPG---AGTTPA---PGY-PAQGYTGAQPG 79
QY 77 ---VYRGPSPG-----PGAYPSSGQSPAPG-AYATGPF-----GAPAGPLIVPNLPL 120

Db 80 GAPSYGVPPGGFVGGAGFSGYPPPSQSYSGGAPVLPGLPGFPGQPPSQY---- 135
QY 121 PGG--VVRMLTITL----GTVKPMAN 141
Db 136 PGGGPTPSQGPATVQVQGITRPMAN 162

RESULT 49

US-08-175-155-69
; Sequence 69, Application US/08175.155
; Patent No. 5641648

GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
APPLICANT: Crisman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: Methods for Preparing Synthetic
TITLE OF INVENTION: Repetitive DNA
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSES: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Rowland, Bettian I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249

INFORMATION FOR SEQ. ID NO. 69:

SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-69

Query Match 14.4%; Score 195.5; DB 1; Length 417;
Best Local Similarity 40.1%; Pred. No. 3.7e-10;
Matches 55; Conservative 8; Mismatches 55; Indels 19; Gaps 7;

QY 9 DALSSGNNPPGQPGAWGNQ----PAGAGYVGAAYPPG-QAPPGA--YVGAAPPA 61
Db 31 DPMGAGTGPQGLPGSPGAPGTGPQGLPGSPA--PCTPGQGLPGSPGAPGTGPQ 88
QY 62 YHGAPEAYGAPAPGVYPPSPGAPYPSGQPSAFGAYAT--GPYGA--GPI 112
Db 89 LRGSPGA-PGTTPQGLPGSPGAPGTGPQGLPGSPGAPGTTPQGLPGSPGAPGTTPQ 147
QY 113 IVPYNLPLPGVVRML 129
Db 148 GLPGSPGAPGTTPQGL 164

RESULT 50
US-08-477-509B-104
; Sequence 104, Application US/08477509B
; Patent No. 5770697

GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A
APPLICANT: Cappello, Joseph
APPLICANT: Crisman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: No. 5770697e1 Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSES: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,509B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986

ATTORNEY/AGENT INFORMATION:

NAME: Treacattin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249

INFORMATION FOR SEQ. ID NO. 104:

SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-509B-104

Query Match 14.4%; Score 195.5; DB 1; Length 417;
Best Local Similarity 40.1%; Pred. No. 3.7e-10;
Matches 55; Conservative 8; Mismatches 55; Indels 19; Gaps 7;

QY 9 DALSSGNNPPGQPGAWGNQ----PAGAGYVGAAYPPG-QAPPGA--YVGAAPPA 61
Db 31 DPMGAGTGPQGLPGSPGAPGTGPQGLPGSPA--PCTPGQGLPGSPGAPGTGPQ 88
QY 62 YHGAPEAYGAPAPGVYPPSPGAPYPSGQPSAFGAYAT--GPYGA--GPI 112
Db 89 LRGSPGA-PGTTPQGLPGSPGAPGTGPQGLPGSPGAPGTTPQGLPGSPGAPGTTPQ 147
QY 113 IVPYNLPLPGVVRML 129
Db 148 GLPGSPGAPGTTPQGL 164

Search completed: August 23, 2004, 14:39:25
Job time : 24 secs